

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ

Katedra softwarového inženýrství

Studijní program: Aplikace informatiky v přírodních vědách



**Analýza vlivu investorovy pozornosti na
ceny finančních aktiv prostřednictvím
transformerových neuronových sítí**

DIPLOMOVÁ PRÁCE

Vypracoval: Bc. Pavel Ježek

Vedoucí práce: doc. Ing. Quang Van Tran Ph.D.

Rok: 2025

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: Ježek Jméno: Pavel Osobní číslo: 494756
Fakulta/ústav: Fakulta jaderná a fyzikálně inženýrská
Zadávající katedra/ústav: Katedra softwarového inženýrství
Studijní program: Aplikace informatiky v přírodních vědách

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Analýza vlivu investorovy pozornosti na ceny finančních aktiv prostřednictvím transformerových neuronových sítí

Název diplomové práce anglicky:

Analysis of the influence of investor's attention on financial assets' prices using transformer neural networks

Pokyny pro vypracování:

1. Vypracovat rešerší na teorii hlubokých neuronových sítí s hlavním zaměřením na transformery
2. Vypracovat stručný přehled využití transformerů na predikci časových řad
3. Navrhnut postup ke shromáždění a zpracování dat na investorovu pozornost a implementovat ho ve vhodném prostředí
4. Implementovat modely transformerových neuronových sítí na predikci finančního trhu s investorovou pozorností
5. Vyhodnotit jak použité modely, tak získané výsledky a následně je porovnat s výsledky jiných prací

Seznam doporučené literatury:

- 1) VASWANI, Ashish et al. Attention is all you need. Advances in Neural Information Processing Systems 30, 2017.
- 2) WANG, Shuzhen. A stock price prediction method based on BiLSTM and improved transformer. IEEE Access 11, 104211-104223, 2023.
- 3) ZHANG, Qiuyue, et al. Transformer-based attention network for stock movement prediction. Expert Systems with Applications, 2022, 202: 117239.
- 4) YAÑEZ, Camilo; KRISTJANPOLLER, Werner; MINUTOLO, Marcel C. Stock market index prediction using transformer neural network models and frequency decomposition. Neural Computing and Applications 36, 5777–15797, 2024.
- 5) HARYONO, Agus Tri; SARNO, Rianarto; SUNGKONO, Kelly Rossa. Transformer-gated recurrent unit method for predicting stock price based on news sentiments and technical indicators. IEEE Access 11, 2023.
- 6) BISHOP, Christopher M.; BISHOP, Hugh. Deep learning: Foundations and concepts. Springer Nature, 2023.

Jméno a pracoviště vedoucí(ho) diplomové práce:

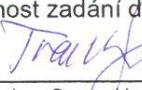
doc. Ing. Quang Van Tran, Ph.D. katedra softwarového inženýrství FJFI

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

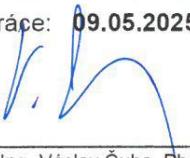
Datum zadání diplomové práce: **29.10.2024**

Termín odevzdání diplomové práce: **09.05.2025**

Platnost zadání diplomové práce: **28.10.2026**


doc. Ing. Quang Van Tran, Ph.D.
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry


doc. Ing. Václav Čuba, Ph.D.
podpis děkana(ky)

PROHLÁŠENÍ

Já, níže podepsaný(á)

Příjmení, jméno studenta: Ježek, Pavel

Osobní číslo: 494756

Název programu: Aplikace informatiky v přírodních vědách

prohlašuji, že jsem bakalářskou/diplomovou práci s názvem

**Analýza vlivu investorovy pozornosti na ceny finančních aktiv prostřednictvím
transformerových neuronových sítí**

vypracoval(a) samostatně a uvedl(a) veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací a Rámcovými pravidly používání umělé inteligence na ČVUT pro studijní a pedagogické účely v Bc a NM studiu.

Prohlašuji, že jsem v průběhu příprav a psaní závěrečné práce použil/a nástroje umělé inteligence. Vygenerovaný obsah jsem ověřil/a. Stvrzuji, že jsem si vědom/a, že za obsah závěrečné práce plně zodpovídám.

Prohlašuji, že jsem v průběhu příprav a psaní závěrečné práce nepoužil/a žádný nástroj umělé inteligence. Jsem si vědom/a důsledků, kdy bude zjevné nepřiznané použití těchto nástrojů při tvorbě jakékoli části mé závěrečné práce.

v Táboré dne 3.8.2025


Podpis

Poděkování

Děkuji vedoucímu doc. Ing. Quang Van Tranovi Ph.D. za neocenitelné rady a pomoc při tvorbě diplomové práce.

Bc. Pavel Ježek

Název práce:

Analýza vlivu investorovy pozornosti na ceny finančních aktiv prostřednictvím transformerových neuronových sítí

Autor: Bc. Pavel Ježek

Studijní program: Aplikace informatiky v přírodních vědách

Druh práce: Diplomová práce

Vedoucí práce: doc. Ing. Quang Van Tranovi Ph.D.

Katedra softwarového inženýrství, Fakulta jaderná a fyzikálně inženýrská, České vysoké učení technické v Praze

Abstrakt: Tato práce se zabývá využitím transformerových modelů pro predikci cen a trendů na finančních trzích, a to s důrazem na zapojení tzv. investorovy pozornosti. Investorovu pozornost zde reprezentují nejen tradiční finanční data, ale také informace o náladě ve společnosti, získané prostřednictvím analýzy sentimentu z různých veřejně dostupných zdrojů, jako jsou vyhledávací trendy, články v médiích či příspěvky na sociálních sítích. Cílem práce je ukázat, že propojení tradičních a netradičních datových vrstev může přinést přidanou hodnotu při modelování dynamiky finančních trhů a že transformerové modely představují flexibilní a výkonný nástroj pro takové úlohy.

Klíčová slova: Transformer, Investorova pozornost, finance, akciový trh, predikce na akciovém trhu

Title:

Analysis of the influence of investor's attention on financial assets' prices using transfromer neural networks

Author: Bc. Pavel Ježek

Abstract: This thesis focuses on the use of transformer models for predicting prices and trends in financial markets, with an emphasis on the involvement of so-called investor attention. Investor attention is represented not only by traditional financial data, but also by information about the mood in society, obtained through sentiment analysis from various publicly available sources, such as search trends, media articles, and social media posts. The aim of this work is to show that combining traditional and non-traditional data layers can bring added value to the modeling of financial market dynamics and that transformer models are a flexible and powerful tool for such tasks.

Key words: Transformer, Investor's attention, finance, stock market, prediction on stock market

Obsah

Seznam zkratek.....	8
Úvod	10
1 Hluboké učení	12
1.1 Neuronová síť	12
1.2 Architektury.....	13
1.2.1 Rekurentní neuronové sítě	13
1.2.2 Konvoluční sítě	15
1.3 Motivace vzniku Transformerů	16
2 Transformer a jeho architektura	17
2.1 Mechanismus pozornosti	17
2.1.1 <i>Scaled dot-product attention</i>	17
2.1.2 <i>Self-attention a cross-attention</i>	18
2.1.3 <i>Multi-head attention</i>	18
2.1.4 <i>Multi-query attention a grouped query attention</i>	19
2.2 Architektura modelu	19
2.2.1 Vnoření slov a poziční kódování	20
2.2.2 Enkodér a Dekodér	20
2.3 Strategie trénování	21
2.3.1 <i>Masked language modeling</i>	21
2.3.2 <i>Causal language modeling</i>	21
2.3.3 <i>Sequence-to-sequence</i>	21
2.4 Ztrátové funkce.....	22
2.4.1 Křížová entropie.....	22
2.4.2 Střední kvadratická chyba.....	22
2.4.3 Střední absolutní chyba.....	22
2.4.4 Huberova ztrátová funkce	22
2.5 Optimalizační techniky	23
2.5.1 Optimalizátory.....	23
2.5.2 Rychlosť učení.....	23
2.5.3 Regularizace.....	24
2.5.4 Stabilizační techniky	24
2.6 Varianty modelů	25
2.6.1 <i>BERT</i>	25
2.6.2 <i>GPT</i>	26

2.7	Využitý model.....	27
2.7.1	<i>Temporal Fusion Transformer</i>	27
2.7.2	Zjednodušená verze	30
3	Rešerše současného využití ve finančním sektoru.....	31
3.1	Sentimentální analýza	31
3.2	Detekce podvodů a anomálií.....	31
3.3	Modelování časových řad.....	32
3.4	Předpovědi časových řad.....	32
4	Data	34
4.1	Finanční trhy.....	34
4.1.1	Apple	34
4.1.2	Bitcoin.....	37
4.1.3	Zlato.....	38
4.2	Technické indikátory a zpožděné proměnné	40
4.2.1	Klouzavé průměry.....	40
4.2.2	Bollingerova pásmá	41
4.2.3	Moving Average Convergence Divergence (MACD)	42
4.2.4	Index Relativní síly	42
4.2.5	Fourierovy transformace	43
4.3	Investorova pozornost.....	44
4.3.1	Trendy vyhledávání.....	44
4.3.2	Reddit	45
4.3.3	Novinové titulky	49
5	Modely a jejich výsledky	50
5.1	Predikce trendu	50
5.1.1	Apple	50
5.1.2	Bitcoin.....	50
5.1.3	Zlato.....	51
5.2	Predikce ceny.....	51
5.2.1	Apple	51
5.2.2	Bitcoin.....	54
5.2.3	Zlato.....	56
5.3	Porovnání s jinými modely	58
	Závěr	61
	Zdroje.....	63

Seznam zkratek

Adagrad	Adaptive Gradient Algorithm	Adaptivní algoritmus gradientu
Adam	Adaptive Moment Estimation	Adaptivní odhad momentů
ANOVA	Analysis of Variance	Analýza rozptylu
BERT	Bidirectional Encoder Representations from Transformers	Dvousměrné kódování pomocí transformátorů
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise	Ensemble dekompozice s adaptivním šumem
CLM	Causal Language Modeling	Kauzální jazykové modelování
CNN	Convolutional Neural Network	Konvoluční neuronová síť
DADELTA	Delta Accumulated Gradient (AdaDelta)	Akumulovaný delta gradient (AdaDelta)
EMA	Exponential Moving Average	Exponenciální klouzavý průměr
EMD	Empirical Mode Decomposition	Empirická dekompozice módu
FinBERT	Financial BERT	BERT model trénovaný na finančních textech
FTSE 100	Financial Times Stock Exchange 100	Index 100 největších firem v UK
GAN	Generative Adversarial Networks	Generativní nepřátelská síť
GNN	Graph Neural Network	Grafová neuronová síť
GPT	Generative Pre-trained Transformer	Generativní předtrénovaný transformátor
GRN	Gated Residual Network	Bránová reziduální síť
GRU	Gated Recurrent Unit	Bránová rekurentní jednotka
HSI	Hang Seng Index	Akciový index v Hongkongu
Informer	Informer (efficient time-series model)	Efektivní model pro časové řady
Leaky ReLU	Leaky Rectified Linear Unit	Netěsná usměrněná lineární jednotka
LlaMA	Large Language Model Meta AI	Velký jazykový model od Meta AI
LLM	Large Language Model	Velký jazykový model
LoRA	Low-Rank Adaptation	Nízkodimenzionální adaptace
LOTSA	Long-Term Sequence Alignment	Zarovnání dlouhých sekvencí
LSTM	Long Short-Term Memory	Paměť s dlouhým a krátkým dosahem
MACD	Moving Average Convergence Divergence	Konvergence/divergence klouzavých průměrů
MAE	Mean Absolute Error	Průměrná absolutní chyba
MAPE	Mean Absolute Percentage Error	Průměrná absolutní procentní chyba
MLM	Masked Language Modeling	Maskované jazykové modelování
MLP	Multi-Layer Perceptron	Vícevrstvý perceptron
MOIRAI	Multimodal Forecasting Architecture	Multimodální architektura pro predikci
MSE	Mean Squared Error	Průměrná kvadratická chyba
NASDAQ	Nasdaq Stock Market	Technologický akciový trh USA
NIFTY50	Nifty 50	Index 50 největších firem v Indii
Nikkei 225	Nikkei 225	Japonský akciový index
NLP	Natural Language Processing	Zpracování přirozeného jazyka
NSP	Next Sentence Prediction	Predikce následující věty
PATHformer	Path-based Transformer	Transformátor založený na cestách (v grafech/datových tocích)

ReLU	Rectified Linear Unit	Usměrněná lineární jednotka
RFE	Recursive Feature Elimination	Rekurzivní eliminace příznaků
RMSE	Root Mean Squared Error	Střední kvadratická odchylka
RMSprop	Root Mean Square Propagation	Šíření střední kvadratické hodnoty
RNN	Recurrent Neural Network	Rekurentní neuronová síť
RoBERTa	Robustly optimized BERT approach	Robustně optimalizovaný přístup BERT
S&P 500	Standard & Poor's 500	Index 500 největších amerických společností
SDTO	Symbolic Data Trend Operator	Symbolický operátor trendu dat
SGD	Stochastic Gradient Descent	Stochastický gradientní sestup
SHAP	SHapley Additive exPlanations	SHapleyho aditivní vysvětlení
SMA	Simple Moving Average	Jednoduchý klouzavý průměr
STOXX50	Euro STOXX 50	Index 50 největších evropských společností
TEANet	Temporal Event Attention Network	Síť s pozorností na časové události
TFT	Temporal Fusion Transformer	Časově sloučený transformátor
TIMER	Time Series Transformer	Transformátor pro časové řady
TIMEXER	Transformer for Time Series Forecasting with Exogenous Variables	Transformátor pro predikci časových řad s externími proměnnými
TRABSA	Transformer-Based Sentiment Analysis	Analýza sentimentu založená na transformerech
VTT	Vision-to-Text	Převod vizuální informace na text
XLex	eXplainable Lexicons	Vysvětlitelné lexikony (slovníky pro XAI a NLP)

Úvod

Předpověď cen finančních aktiv představuje dlouhodobě jeden z nejkomplexnějších a zároveň nejzádanějších problémů v oblasti ekonomie, financí i datové vědy. Jedná se o obtížný úkol, především kvůli těžké předvídatelnosti trhů, které jsou ovlivňovány širokým spektrem faktorů – od makroekonomických ukazatelů přes geopolitické události až po sentiment investorů. Finanční časové řady často vykazují vysokou volatilitu, nestacionaritu a nelineární chování, což ztěžuje aplikaci tradičních statistických metod. Navíc dochází k častým změnám tržních režimů, kdy modely dobře fungující v určitém období selhávají v jiném. Dalším významným faktorem je asymetrická informovanost účastníků trhu, která vede k iracionálním rozhodnutím.

V posledních letech se proto čím dál více pozornosti obrací k behaviorálním aspektům finančních trhů, zejména k tomu, jak informace vnímají investoři. Tím se zabývá investorova pozornost. Přesněji tento koncept představuje míru toho, jak výrazně si investoři všímají určitých témat, událostí nebo aktiv. Změny v této pozornosti mohou mít významný dopad na rozhodování investorů a tím i na samotné pohyby cen na trzích. S rozvojem digitálních technologií, sociálních sítí a medií se navíc objevují nové možnosti, jak tuto pozornost kvantifikovat.

Tato práce si klade za cíl analyzovat, využít informace o investorově pozornosti pro predikci cen finančních aktiv. K tomu využívá moderní přístupy strojového učení, konkrétně transformerové neuronové sítě, které se ukázaly jako mimořádně efektivní při zpracování sekvenčních a jazykových dat. Díky své schopnosti zachytit dlouhodobé závislosti a kontextová spojení mají Transformery potenciál lépe porozumět vztahu mezi informačním prostředím a tržními pohyby cen než tradiční modely.

Vybraný přístup je v této práci prezentován na 3 různých druzích finančních komodit. Jedná se o akcie (Apple), kryptoměnu (Bitcoin) a opci (zlato). Práce je rozdělena do 5 kapitol.

První kapitola popisuje oblast hlubokého učení. Zabývá se jeho původem a principem vzniku. Nejdříve přiblížuje základní strukturu neuronů a neuronových sítí. Pokračuje popisem rozšířených architektur, jmenovitě se jedná o rekurentní neuronové sítě a konvoluční neuronové sítě. U rekurentních sítí se zaměřuje na jejich specifické vlastnosti a typické modely použití. U konvolučních popisuje jejich strukturu a operaci konvoluce na datech.

Druhá kapitola pojednává o Transformerech. Nejprve objasňuje klíčový prvek transformerových modelů – mechanismus pozornosti. Popisuje základní myšlenky a představuje různé druhy tohoto mechanismu. Pokračuje přehledem architektury modelu, z jakých částí se skládá a jaká je jejich funkce. Další podkapitola se zabývá různými cíli a způsoby tréninku, které se využívají pro tvorbu modelů. S tím souvisí navazující část s popisem některých ztrátových funkcí. Spojitost s tréninkem mají i optimalizační metody, popsané v další podkapitole. Kapitolu zakončuje ukázka dnes nejrozšířenějších modelů a popis modelu použitého pro samotné predikce.

Třetí kapitola je zaměřena na rešerši současného využití, převážně ve finančním sektoru. Je rozdělena do několika podkapitol podle zaměření jednotlivých prací, od sentimentální analýzy až po předpověď či modelování časových řad.

Čtvrtá kapitola se věnuje použitým datům. Rozdělena je do 3 hlavních částí na základě jejich původu. Nejdříve je popis finančních dat, která slouží jako výchozí bod. Na jejich základě pokračuje odvození zpožděných proměnných a jiných indikátorů. Poslední podkapitola se

zabývá sentimentální analýzou, která představuje další vrstvu informací odrážející pozornost a náladu investorů.

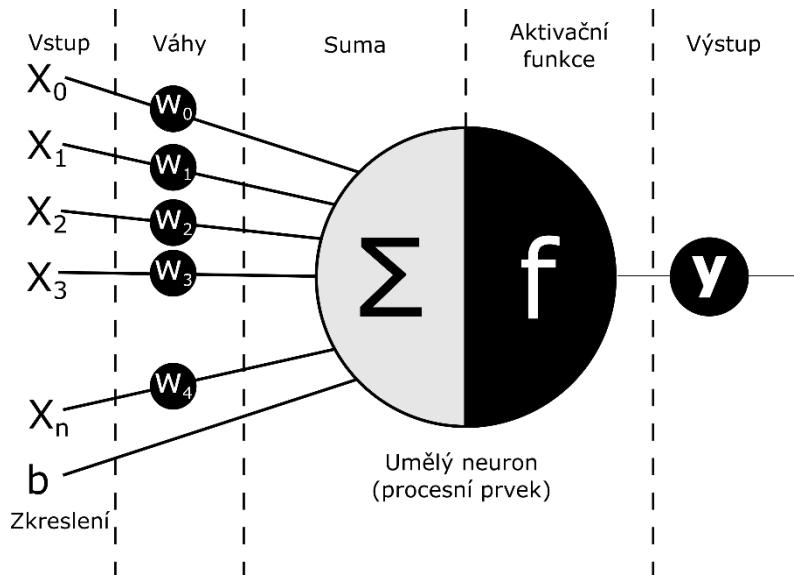
Pátá, závěrečná kapitola je o vykreslení vlastností použitých modelů. Popisuje rozsah využitých dat, zvolené hyperparametry a prezentuje výsledky pomocí různých metrik a grafů. Výstupy jsou tvořeny pro dva typy úloh – predikci trendu a předpověď ceny následujícího dne. Kapitolu uzavírá porovnání s dalšími současnými pracemi na téma predikcí na finančních trzích, speciálně ceny akcií Apple.

1 Hluboké učení

Hluboké učení je jednou z nejvíce využívaných metod strojového učení současnosti. Využívá modely inspirované chováním neuronů v mozku nazvané neuronové sítě. Vrstvením těchto sítí jsou vytvořeny modely navržené pro různé úkoly, jako jsou například klasifikace nebo regrese (problémy, které jsou významně spojovány s ostatními odvětvími strojového učení). Tyto modely se aplikují na problémy v kategorii učení s učitelem, učení bez učitele, tak i jejich kombinaci (původním názvem *semi-supervised*, volně lze přeložit jako učení s učitelem i bez učitele).

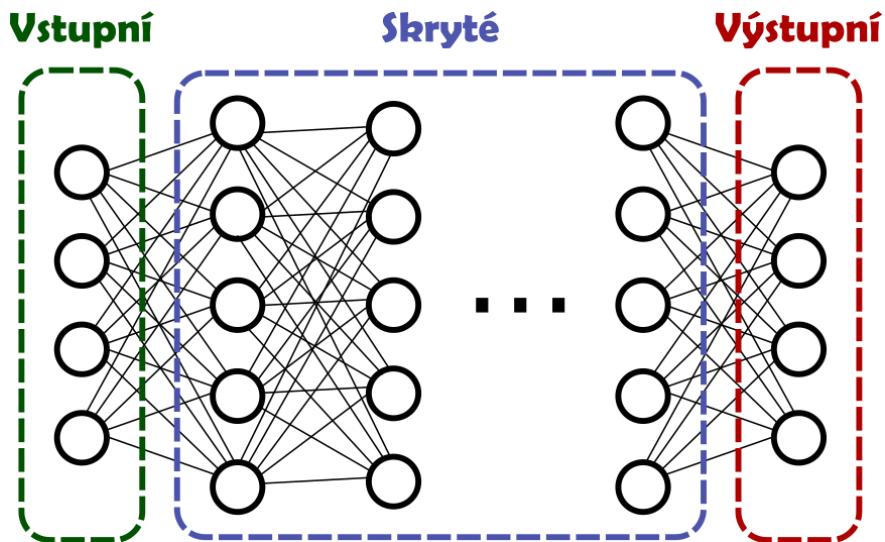
1.1 Neuronová síť

Základním prvkem neuronových sítí je neuron. Jak už bylo zmíněno výše, tento prvek si bere za vzor stejně jmenovaný druh buňky nacházející se v mozku. Vazby mezi jednotlivými neurony se nazývají synapse (pomyslně je lze rozdělit stejně jako u buněčné struktury na dendrity a axon – vstupy a výstup). Základní schéma lze vidět na obrázku 1. Na něm lze vypozorovat i cestu jednotlivých informací. Na vstupu je vektor představující informaci (případně je k němu připojeno i zkreslení – *bias*). Ten prochází synapsem (dendrity), na kterých jsou na jednotlivé jeho části aplikovány váhy. Dále informace vchází do samotného neuronu, kde dochází k využití sumační a aktivační funkce. Výstup pak přechází do dalšího neuronu a situace se opakuje (nebo se jedná o poslední vrstvu a jde o výstup části či celého modelu) [1].



Obrázek 1: Grafické znázornění neuronu a cesty informace skrz něj, upraveno a převzato z [1]

Síť samotná je tvořena neurony uskupenými do skupin podle synapsí – vrstev. Její skladba se rozděluje do tří hlavních částí (ty lze vidět na obrázku 2). Na začátku data putují do vstupní vrstvy. Jedná se o viditelnou část sítě, jelikož její jednotlivé proměnné jsou předem dané. Dále se pokračuje do skrytých vrstev. Tato část je složena z N neuronových vrstev, kde N je celé číslo větší než 2 (pro hodnotu 1 se pak jedná o *MLP*, ta se většinou řadí spíše do strojového učení než hlubokého učení). Z názvu již vyplývá, že u těchto vrstev nelze nahlédnout na vnitřní parametry, model si je určuje sám. Jejich cílem je informace z dat zjemnit na abstraktnější vlastnosti. Poslední částí je výstupní vrstva, určená pro převedení do řešení daného problému. U této vrstvy pak závisí na typu dat (množina označení plynoucí z této vrstvy je spojité, binární nebo pravděpodobnostní rozdělení nad třídami) [1].



Obrázek 2: Znázornění vrstev v síti

1.2 Architektury

Neuronové sítě lze dělit podle jejich architektury, která se odvíjí od specifických požadavků různých problémů či dat. Existuje velké množství různých typů, podrobněji se zaměříme na ty nejpoužívanější. Mezi ně patří rekurentní neuronové sítě (*RNN*) a konvoluční neuronové sítě (*CNN*).

1.2.1 Rekurentní neuronové sítě

Tento typ sítí se používá pro sekvenčně závislá data, jako jsou časové řady nebo texty. V praxi se nejčastěji využívá pro symbolické hodnoty. Jednotlivé vrstvy v *RNN* korespondují s pozicemi dat v sekvenci, proto mají proměnlivý počet skrytých vrstev. Díky tomu se zachovává i informace závislá na pořadí, protože následující prvek nese lehce jinou informaci na základě prvku, který mu předchází. Stav, kdy neuron bere v potaz (v biologické terminologii lze hovořit o „vzpomínání“) předcházející data, se nazývá *hidden state* (skrytý stav) nebo také *hidden memory*. Tyto sítě mají také vlastnost tzv. Turingovské kompletnosti, což znamená, že lze jimi simulovat jakýkoliv algoritmus (za předpokladu dostatečných výpočetních zdrojů). Na základě tohoto přístupu vznikají různé varianty. Obecně se v praxi nejčastěji setkáváme s *LSTM* nebo *GRU* [2] [3].

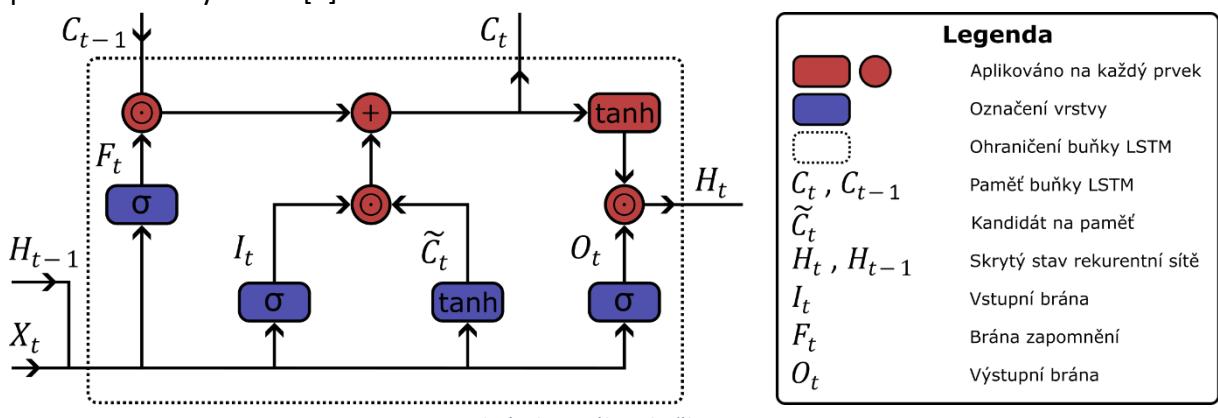
Long Short-Term Memory neboli *LSTM* je návrh, který vznikl pro řešení problémů vznikajících při učení základních modelů *RNN*. Zejména se jedná o problém mizejícího nebo explodujícího gradientu. Ty znemožňují efektivní učení závislostí v dlouhých sekvencích, a tedy odhalit kontext v dlouhých posloupnostech, jako jsou finanční časové řady či textová data.

Proto *LSTM* modely zavádějí koncept paměťových buněk. Jejich hlavní funkcí je uchovat informace po delší časový úsek. Jejich obsluhou se zabývá sada řídících bran podobných logickým branám v informatice.

Jeho základními částmi jsou:

- Vstupní brána
- Tato brána rozhoduje, společně s kandidátem na paměť, jaká nová informace bude přidána do paměti buňky. Kontrolou přidávání nových informací se omezuje šum a zlepšuje stabilitu učení.
- Brána zapomnění
- Tato část ovládá množství paměti, která bude zapomenuta. Jedná se o specifickou vlastnost, kterou nemají klasické *RNN*. Tím se zabraňuje nahromadění nerelevantních dat a přehlcení paměti buňky. Jejím základním principem je selektivní čištění paměťového stavu a díky tomu zachovává pouze informace důležité pro aktuální kontext.
- Výstupní brána
- Poslední brána určuje, jaká část paměti buňky se použije pro výpočet skrytého stavu. Díky tomu lze ovládat tvar výstupu na základě aktuální i minulé informace.

Celý princip lze vidět na obrázku 3. Tyto mechanismy zaručují, že model *LSTM* efektivně modeluje dlouhodobé závislosti. Pro tuto vlastnost jsou vyhledávanou architekturou pro predikce časových řad [3].



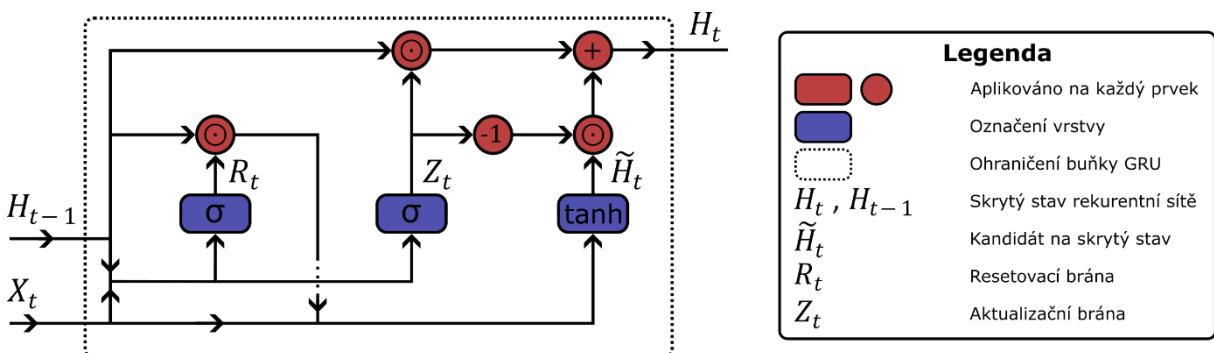
Obrázek 3: Nákres buňky LSTM

Druhý zmíněným modelem je *Gated Recurrent Unit* neboli *GRU*. Jedná se o novější zjednodušenou variantu výše zmíněné *LSTM* a samozřejmě s ním sdílí i hlavní cíl, řešit problémy *RNN* pomocí zachovávání informací na delší časové vzdálenosti. Na rozdíl výše od výše zmíněné varianty neobsahuje vlastní paměťovou buňku, kvůli čemuž má pouze dvě brány.

Konkrétně se skládá z:

- Resetovací brána
 - Tato složka ovlivňuje, kolik z předchozího skrytého stavu se použije pro výpočet její kandidátní hodnoty (ekvivalent kandidáta v *LSTM*). Pokud brána vyšle malou hodnotu, většina minulého stavu se zapomene a naopak. Toto zapomínání je vhodné pro zpracovávání krátkodobých závislostí, jedná se o výběr s jak hlubokou minulostí daný krok pracuje.
- Aktualizační brána
 - Tato brána určuje, jaká část z kandidáta se začlení do aktuálního skrytého stavu (tím pádem i kolik bude zachováno z předešlého). V podstatě spojuje zapomínání a ukládání z *LSTM*. Modelu to umožňuje stabilnější přenášení informace napříč delšími sekvencemi ale i adaptovat se na nové podněty

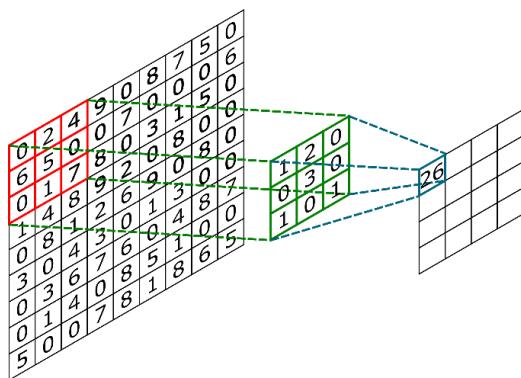
Jednodušší struktura *GRU* napomáhá rychlejšímu učení a zároveň je méně výpočetně náročná. To vše bez výrazné ztráty přesnosti. Proto je vhodná na rozsáhlá data [3].



Obrázek 4: Nákres buňky *GRU*

1.2.2 Konvoluční sítě

Druhým zmíněným zástupcem jsou konvoluční sítě. Ty jsou navrženy pro práci s daty v mřížce, čímž zachovávají informace i o jejich poloze v datech (např. u obrázku se uchovávají informace o sousedních pixelech, které by se transformací do nižší dimenze ztratily). Základní operací, jak už z názvu vyplývá, jsou konvoluce. Jedná se o matematickou operaci, která kombinuje 2 funkce a vytváří novou s vlastnostmi původních. V rámci neuronových sítí se pak jedná o aplikaci filtru (jádra) na část dat, graficky znázorněno na obrázku 5. Filtr se pak pohybuje v rámci celé mřížky a v každém kroku udělá bodový součin. Výsledkem je pak mapa znaků dat, která zdůrazňuje vlastnosti dat na základě použitého filtru. Tato operace se pak sdružuje v části nazvané konvoluční vrstva [3].



Obrázek 5: Jednoduchý grafický popis operace konvoluce

Obecně se poté vrstvy dají shrnout do základní sady následovně:

- Vstupní vrstva
- Konvoluční vrstva
- Aktivační vrstva
 - Nelineární transformace, nejčastěji použitím funkce *ReLU* nebo *leaky ReLU*
- Sdružovací a vzorkovací vrstva
 - Snížení rozlišení mapy znaků
- Normalizační vrstva
- Vrstva odpadnutí (*dropout*)
 - Odstranění části neuronů z předchozí vrstvy
- Plně propojená vrstva

1.3 Motivace vzniku Transformerů

Ostatní modely hlubokého učení mají několik omezení, které mohou být částečně překonány pomocí architektury Transformerů. Pro modely hlubokého učení s vysokou přesností je potřeba velké množství dat. Transformery díky svým mechanismům, jako je *self-attention* (sebepozornost), efektivněji pracují s dostupnými daty a lépe zobecňují i na menších datasetech. Zároveň se předtrénované modely dají pomocí malého množství dat pomocí *fine-tuning* metody přeměnit pro využití na specifický problém (*GPT* a *BERT* modely).

Specificky *RNN* modely mají problémy se zpracováváním dlouhých posloupností dat (problém mizejících gradientů). *Self-attention* mechanismus u Transformerů ale dokáže odlišně vážit části dat bez ohledu na jejich vzdálenost a díky tomu zpracuje i dlouhé sekvence informací. Zároveň lze zpracovávat data paralelně, což výrazně urychluje trénování oproti rekurentním modelům pracujícím sekvenčně [4].

2 Transformer a jeho architektura

Transformery představují revoluční architekturu v oblasti hlubokého učení, která významně posunula možnosti zpracování přirozeného jazyka (NLP) a dalších sekvenčních dat. Poprvé byla představena v roce 2017 v článku *Attention Is All You Need* od Vaswaniho a kolegů [4]. Staly se základem pro mnoho moderních aplikací umělé inteligence, speciálně v oblasti strojového překladu, generování textu a analýzy sentimentu.

2.1 Mechanismus pozornosti

Jedná se o základ architektury Transformerů. První zmínku popisu tohoto mechanismu lze nalézt v práci *Neural machine translation by jointly learning to align and translate* od Bahdanau a kolegů [5]. V základní podobě lze definovat jako:

$$\text{Attention}(q, D) = \sum_{i=1}^m \alpha(q, k_i) v_i, \quad (1)$$

kde q označuje query (dotazy), k_i je klíč pro dvojici i v databázi, v_i je hodnota pro dvojici i v databázi, D pak označuje samotnou databázi definovanou jako $D = \{(k_1, v_1), \dots, (k_m, v_m)\}$, m je počet dvojic v databázi a $\alpha(q, k_i)$ označuje váhy pozornosti [3] [5]. Hodnoty vah pak udávají relevanci pozornosti (podobnosti) jednotlivých částí databáze na sekvenci. Ty vznikají jako normalizované hodnoty podobnosti mezi dotazem a klíčem (v případě práce od Bahdanau se tato hodnota počítá pomocí malé neuronové sítě). Speciálně se pak využívá *softmax* funkce přes všechny prvky. Celková operace výpočtu pozornosti se také nazývá *attention pooling* neboli sdružování pozornosti.

2.1.1 Scaled dot-product attention

Tým Vaswaniho [4] využívá již zmíněnou notaci s přidáním velikosti vektoru klíčů a změnou získávání hodnoty podobnosti, která nevyužívá neuronové sítě. Nově vzniklou formulaci nazvali *scaled dot-product attention* neboli ve volném překladu se jedná o pozornost využitím škálovaného skalárního součinu. Matematická formulace má tvar:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (2)$$

kde Q je matice více queries (dotazů), K je matice klíčů, V je matice hodnot a d_k je dimenze klíčů. Jedná se o úpravu *Dot-Product Attention*, která by měla pro velké hodnoty dimenze klíčů zabránit *softmax* funkci dávat hodnoty s malými gradienty. Proto také název *scaled*, pro upřesnění změny (škálování) velikosti výsledné pozornosti [4].

2.1.2 Self-attention a cross-attention

Self-attention označuje mechanismus, který dynamicky váží význam jednotlivých částí vstupní sekvence vzhledem k sobě samým. Tím definuje, že matice dotazů, klíčů i hodnot pochází z jedné matice vstupních dat (jedná se o princip užívání vstupních hodnot, nemění samotný způsob výpočtu). Matice jednotlivých částí vzniknou lineární projekcí vstupních dat pomocí váhových matic:

$$Q = XW^Q, K = XW^K, V = XW^V, \quad (3)$$

kde X označuje původní data a W^Q, W^K, W^V jsou váhové matice pro lineární projekci dotazů, klíčů a hodnot. Ty už je pak možné použít pro výpočet samotné pozornosti. Obecně jsou ve výpočtech popsány situace kdy Q, K a V jsou matice oddělené – nepochází ze stejných dat. Tento přístup je znám pod pojmem *cross-attention*. Lze ji vidět v dekodéru, ve střední vrstvě [3] [4].

2.1.3 Multi-head attention

Z dat chceme zachytit co nejvíce informací, například závislosti délky na sekvenci. Vylepšení pak lze docílit rozložením na více podprostorů s rozdílnými lineárními projekcemi (matice vah) a na každém provést pozornostní sdružování. Ty jsou pak sloučeny a využitím další projekce vznikne výsledná pozornost zvaná *multi-head attention*, ve volném překladu mechanismu pozornosti využívající více hlav. Pokud si výpočet pozornosti pro každou část (hlavu) definujeme jako:

$$h_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (4)$$

kde Q, K a V a jsou matice dotazů, klíčů a hodnot, matice W_i^Q, W_i^K a W_i^V jsou matice lineárních projekcí pro dotazy, klíče a hodnoty pro jednotlivé hlavy a i označuje index hlavy. Díky tomu lze celkovou pozornost napsat následovně:

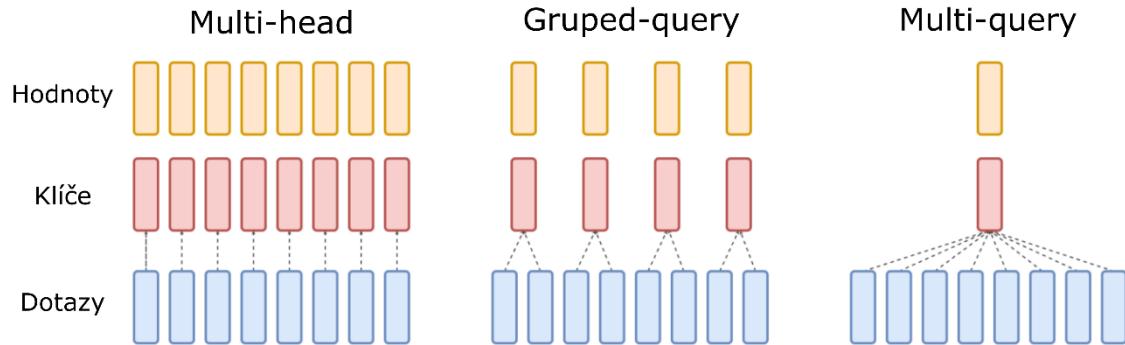
$$\text{MultiHeadAttention}(Q, K, V) = \text{concat}(h_1, \dots, h_n)W^O, \quad (5)$$

kde W^O je matice lineární transformace a concat je funkce pro sjednocení výsledků jednotlivých hlav.

Existuje speciální varianta zvaná *masked multi-head attention* s přídomekem *masked*, která je využívána pro omezení nahlížení na informace z následujících prvků v sekvenci. Tohoto chování se docílí použitím masky, která zakryje budoucí prvky a díky tomu model využívá jen minulé a aktuální. Tohoto efektu se docílí aplikováním masky na matice vah pro lineární projekce [3] [4].

2.1.4 Multi-query attention a grouped query attention

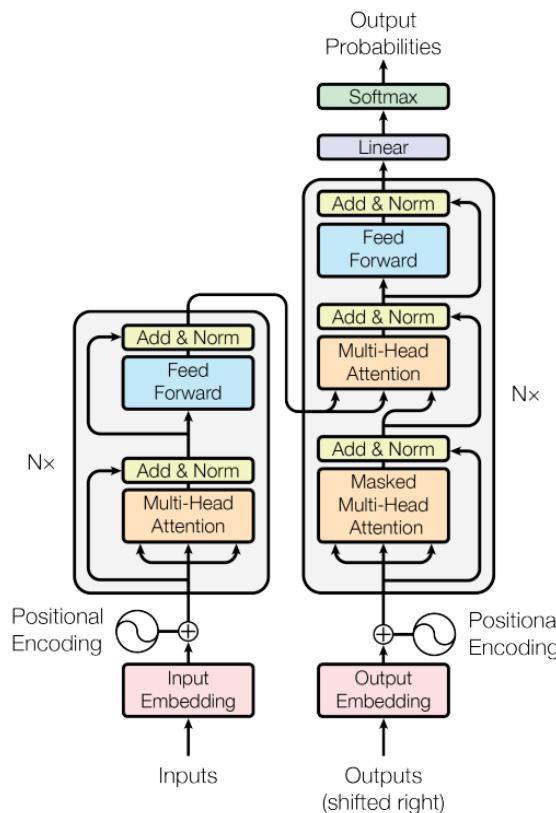
Dnešní modely ale většinou nepoužívají samostatný *multi-head attention*, ale jeho upravené varianty. Jednou z nich je *multi-query attention*. Ta rozděluje na jednotlivé hlavy pouze prostor dotazů, zatímco klíče a hodnoty mají stejné. Další variantou je *Grouped query attention*, která rozděluje vstupní data do skupin. V každé skupině je pak určitý počet hlav, který sdílí prostor klíčů a hodnot ale každá má vlastní dotazy. Na obrázku 6 jsou jejich schematicky zobrazeny [6] [7].



Obrázek 6: Rozdílnosti jednotlivých přístupů mechanismů pozornosti s upraveným textem, původní obrázek pochází z [7]

2.2 Architektura modelu

Základní architektura se skládá ze dvou hlavních částí. Jsou jimi Enkodér a Dekodér. Data před vstupem do těchto částí procházejí úpravou pomocí vnoření slov (*embedding*) a pozičního kódování. Grafické znázornění pohybu informací částmi modelu je vyobrazeno na obrázku 7.



Obrázek 7: Grafické znázornění modelu Transformeru z původní práce, převzato z [4]

2.2.1 Vnoření slov a poziční kódování

Jedná se o první kroky modelů Transformerů. Vnoření slov transformuje diskrétní hodnoty na vektorovou reprezentaci pomocí embedding matice (na začátku náhodná, mění se při trénování). Tento proces je úzce spojen s tzv. tokenizací, podle názvu nově vzniklých vstupů – tokenů. Tokenizace se používá především pro diskrétní data (dnes nejvyužívanější modely jsou pro slova a věty). V případě spojitých dat (například finanční data) se každý krok vezme jako token a transformují se obdobným způsobem (myšleno na vektor užitím matice).

Dále je potřeba předat informaci o pozici v sekvenci, aby model nepracoval s neurčitým balíkem tokenů bez přesných návazností. K tomu se používá poziční kódování. To může být proměnné v rámci trénovací fáze anebo fixně dané na začátku. Základní tvar (fixní) je definován:

$$p(i, 2j) = \sin\left(\frac{i}{10000^{\frac{2j}{d}}}\right), \quad (6)$$

$$p(i, 2j + 1) = \cos\left(\frac{i}{10000^{\frac{2j}{d}}}\right), \quad (7)$$

kde i je pozice (krok při spojitých datech), j je dimenze a d označuje velikost vnoření. Hlavní cílem těchto kroků je připravit data do formy vhodné pro efektivní zpracování modelem. Zároveň však zachovávají význam prvků v sekvenci. Jelikož Transformery nepracují s daty sekvenčně jako *LSTM* či *GRU* a jedná se tedy o klíčové úpravy. [3] [4].

2.2.2 Enkodér a Dekodér

Jak již bylo zmíněno výše, základními částmi Transformerů jsou Enkodér a Dekodér. Enkodér se skládá ze N stejných vrstev, každá obsahuje další 2 podvrstvy. Jedná se o *multi-head attention* vrstvu následovanou plně propojenou vrstvou. Každá z podvrstev je obalena zbytkovým připojením (*residual connection*) a dále vrstvou normalizace. Matematicky lze výstup z podvrstvy vyjádřit jako:

$$\text{LayerNorm}(x + \text{Sublayer}(x)), \quad (8)$$

kde $\text{Sublayer}(x)$ je transformace podvrstvy a $\text{LayerNorm}(\dots)$ je normalizační vrstva. Toto uspořádání umožnuje efektivně zpracovávat celou sekvenci paralelně, a to i se zachováním informace o vztazích mezi prvky. Token získává informaci o ostatních pozicích na vstupu skrze *multi-head attention* vrstvu, která si dynamicky určuje, na kterou se „chce“ zaměřit. Plně propojená vrstva přidává nelinearitu a zlepšuje tím učení složitějších vzorů. Reziduální připojení s normalizací zaručuje určitou stabilitu učení a také jeho efektivitu.

Dekodér je také složený z více vrstev, avšak každá z nich má 3 podvrstvy. Cílem této části modelu je generovat sekvence autoregresivním způsobem (predikce se opírají pouze o minulé hodnoty společně s výstupy z Enkodéru). To zajišťuje kombinace *masked multi-head attention* vrstvy a *multi-head attention* vrstvou spojující výstupy z enkodérů (zastupující klíče a hodnoty) a data procházející dekodérem (dotazy). Spojení z druhé podvrstvy umožňuje se Dekodéru zaměřit na relevantní části vstupu. Vrstvu zakončuje plně propojená vrstva.

Všechny plně propojené vrstvy mají tu vlastnost, že se aplikují na každou pozici zvlášť, avšak zároveň se stejnými parametry. Jedná se o dvě po sobě jdoucí lineární transformace, mezi kterými leží aktivační *ReLU* funkce [4].

2.3 Strategie trénování

Způsobů hodnocení modelů při tréninku je mnoho. Ty se liší jak ve strategii nakládání s daty, tak v užitých funkčích ztráty. V této části budou vyjmenovány vybrané způsoby hojně využívané v jiných modelech.

2.3.1 *Masked language modeling*

Jedná se o předtrénovací strategii v současnosti užívanou u modelů typu *BERT*. V rámci trénování je část vstupních tokenů náhodně zakryta maskou a model má za úkol predikovat jejich původní hodnotu. Tu získává pouze z kontextu okolních hodnot. Tím je model nucen naučit se vztahy mezi daty (slovy v tomto případě) v obou směrech. Koncept je zaměřen na práci s jazykem, ale lze ho upravit i na jiné varianty dat, jako jsou například časové řady. Výhodou této metody je její samoučící charakter, protože nepotřebuje anotovaná data. Cílové hodnoty jsou již obsaženy v samotném datasetu a proto je atraktivní pro rozsáhlé datové soubory. Nejčastěji tento přístup pak využívá klasifikační typ ztrátové funkce, typicky křížovou entropii. [8].

2.3.2 *Causal language modeling*

Tento typ trénovací strategie je užíván u autoregresivních modelů, jako jsou *GPT* modely. Ten se oproti *MLM* zaměřuje na vztahy v jednom směru. Nutí model predikovat další znak v sekvenci s použitím informací pouze z předcházejících dat. To se docílí aplikováním kauzální masky pozornosti (ta zaručuje jednosměrnost předávání informací zakrýváním budoucích dat). Opět je nejčastěji jako ztrátová funkce používána křížová entropie [9].

2.3.3 *Sequence-to-sequence*

Lze ji nazvat základním přístupem, jenž využívá plný model prezentovaný v původní vědecké práci (u *MLM* je typicky užíván oboustranný enkodér a u *CLM* pouze struktura s dekodérem). Jeho hlavní účel je řešení strojového překladu, kde vstupní a výstupní data mají jinou velikost. Princip odpovídá popisu základní architektury, kdy Enkodér zpracuje vstup na určitou kontextuální reprezentaci. Tu Dekodér použije pro generaci výstupu. Cílem je předpovědět celou sekvenci na základě vstupní sekvence. Výhodou tohoto přístupu je, že se model učí vnitřní reprezentace závislé pouze na předchozích hodnotách, avšak zohledňují širší kontext vstupní sekvence. Opět je přednostně užívána jako ztrátová funkce křížová entropie, avšak záleží na typu problému. [4].

2.4 Ztrátové funkce

Jedním ze stěžejních prvků při trénování je volba vhodné ztrátové funkce. Ty mají za úkol zjistit, jak dobře daný krok trénování splnil svou úlohu. Jeho volba závisí na typu problému nebo na typu dat.

2.4.1 Křížová entropie

Křížová entropie je jedna z dnes nejvyužívanějších (vyplývá i ze zmínky u *CLM*, *MLM* a *Seq2seq*) funkcí u Transformerů. Tato funkce měří rozdíl mezi skutečným rozdělením pravděpodobnosti a rozdělením daným modelem. Je spojena s pravděpodobnostním modelováním. Má následující tvar:

$$J_E(r, q) = - \sum_{x \in \chi} p(x) \log(q(x)), \quad (9)$$

kde $p(x)$ je diskrétní rozdělení a $q(x)$ je jeho diskrétní approximace. Pro spojité je pak situace obdobná:

$$J_E(r, q) = - \int_{\chi} P(x) \log(Q(x)), \quad (10)$$

kde $P(x)$ a $Q(x)$ jsou spojité varianty z předchozí definice [10].

2.4.2 Střední kvadratická chyba

Střední kvadratická chyba je nejběžnější ztrátovou funkcí u regresních modelů pro spojitá data, jako jsou i finanční. Jedná se o sumu kvadratických chyb jednotlivých pozorování dělenou celkovým počtem prvků. Díky operaci s mocninou se zvýrazňují velké chyby, čímž nutí model je upřednostňovat. I přes nevýhodu citlivosti na extrémní výkyvy je nejen díky své jednoduchosti stále oblíbenou možností pro volbu ztrátové funkce.

2.4.3 Střední absolutní chyba

Další ze standardních přístupu volby funkce je střední absolutní chyba. Ta počítá pouze střední hodnotu chyb (bez aplikace druhé mocniny). Do jednotlivých chyb se nebere v potaz jejich směr. Vlastností je pak, že se disproporčně nevychyluje směrem k velkým chybám. Díky tomu je vhodnější pro datasety s rozdělením s těžšími konci (data s vychýlením na krajích rozdělení, které je větší než exponenciální).

2.4.4 Huberova ztrátová funkce

Jedná se o kompromis mezi střední kvadratickou chybou a střední absolutní chybou. Nabízí velkou diferencovatelnost a zároveň není náchylná k extrémním hodnotám. Funkce se pro malé hodnoty chybá chová jako kvadratická a pro velké jako lineární. Matematicky má tvar:

$$J_H(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2 & \text{pro } |y - f(x)| \leq \delta \\ \delta \cdot \left(|y - f(x)| - \frac{1}{2}\delta \right) & \text{jinak} \end{cases}, \quad (11)$$

kde δ je nastavitelný parametr, y je pravá hodnota a $f(x)$ je hodnota modelovaná. Parametrem δ se nastavuje míra, kterou funkce reaguje na různé vlastnosti dat. Díky tému vylepšeným vlastnostem je vysoce efektivní pro časové řady s občasnými výkyvy do extrémních hodnot [11].

2.5 Optimalizační techniky

Optimalizace u modelů hlubokého učení je důležitý proces, během něhož model upravuje své parametry. To se děje za účelem minimalizace ztrátové funkce. Kvalitní a efektivní optimalizace zaručí správnou generalizaci modelu.

2.5.1 Optimalizátory

Nejvýznamnější úlohou optimalizace je výběr správného optimalizátoru. Existuje více přístupů, dnes jsou nejvyužívanějšími varianty využívající gradientní sestup.

2.5.1.1 Stochastic gradient descent

Jde o základní přístup užívaný u neuronových sítí znám také pod akronymem *SGD*. Formulace je následovná:

$$\theta_{t+1} = \theta_t - \eta \nabla_{\theta} J(\theta; x_i, y_i), \quad (12)$$

kde θ_t je vektor parametrů modelu v kroku t , η je rychlosť učení, J je ztrátová funkce, x_i jsou vstupní data a y_i výstupní. *SGD* pracuje s malými náhodnými vzorky a tím urychluje trénink (tedy i nalezení lokálního optima, v nejlepším případě globálního) [12].

2.5.1.2 Adam (Adaptive moment estimation)

Adam kombinuje výhody adaptivního učení (*Adagrad*) a momentu (*SGD* s momentem). Vypočítává adaptivní míry učení pro každý parametr. Zároveň si z přístupů *Adadelta* a *RMSprop* bere příklad v ukládání exponenciálně ztrácející se průměry kvadratických gradientů. K nim přidává sledování exponenciálně ztrácející se průměrů gradientů jako u momentových typů. Tyto dvě nově vzniklé hodnoty jsou v prvním kroku nulové, a to by mohlo zkreslovat algoritmus (hlavně pokud jsou parametry *Adam* funkce blízko 1). Kvůli tomu se počítají i nezkreslené odhadы těchto momentů. Všechny tyto momenty jsou pak definovány jako:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} J(\theta), \quad (13)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} J(\theta))^2, \quad (14)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (15)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad (16)$$

kde β_1 s β_2 jsou parametry vah minulých hodnot (původní autor doporučuje hodnoty 0,9 a 0,999), m_t a v_t označují první (průměr) a druhý (rozptyl) moment a \hat{m}_t s \hat{v}_t jsou odhadы těchto momentů. Tyto momenty jsou pak užity pro aktualizaci parametrů modelu:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \quad (17)$$

kde η je míra učení a ϵ malá hodnota pro stabilitu dělení [12].

2.5.2 Rychlosť učení

Rychlosť učení určuje velikost kroků podniknutých se při aktualizaci parametrů. Příliš vysoká může způsobit divergenci a nízká zase zpomaluje trénink. Proto se využívají různé schedulery – plánovače, které mění její míru dynamicky.

2.5.2.1 Kosinový annealing

Jedna z oblíbených metod. Snižuje míru učení podle kosinové křivky. Začíná s poměrně velkou hodnotou, rychle klesne na minimum a znova zvedne hodnotu zpátky na maximum. Tím pak dojde k simulování restartu učícího procesu, ale s použitím už lepších vah. Každý krok pak upravuje hodnotu podle:

$$\eta_t = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i) \left(1 + \cos\left(\frac{T_{cur}}{T}\pi\right)\right), \quad (18)$$

kde η_{\min}^i s η_{\max}^i jsou meze míry učení, T je celkový počet epoch, T_{cur} je počet kroků od posledního restartu (pokud je změna v rámci jednoho běhu, je ekvivalentní hodnotě t , tedy kroku výpočtu) [13].

2.5.2.2 Změna po fixním kroku (*Step scheduler*)

Nejjednodušší varianta změny míry učení. Rychlosť se mění skokově po uplynutí určitého počtu kroků o předem danou část. Technika je dobře interpretovatelná a často se využívá například u učení na fixních časových řadách.

2.5.2.3 Rozehřívací fáze (*Warm-up*)

Při tréninku Transformerů se někdy využívá *warm-up* fáze, kde se míra učení na začátku postupně zvedá. To pokračuje, dokud nedojdeme do určitého (daného parametrem) počtu kroků, kde je maximální. To pomáhá stabilizovat začátek učení, kdy jsou váhy ještě nezměněné.

2.5.3 Regularizace

Regularizační techniky zabráňují přeučení. Penalizují příliš složitá řešení a tím podporují větší generalizaci. V této části budou vyjmenovány některé z nich.

2.5.3.1 Odpadnutí (*Dropout*)

Jedná se o náhodné deaktivování/vypínání některých neuronů. Tím se zabrání modelu ve vytvoření závislostí na určitých neuronech a jejich kombinacích. V Transformerech je doporučováno po *self-attention* vrstvách a plně propojených vrstvách. Obvykle jsou hodnoty pravděpodobnosti odpadnutí mezi 0,1 až 0,3.

2.5.3.2 Krácení vah (*Weight Decay*)

Tato regularizace penalizuje velké hodnoty vah. Do ztrátové funkce se tak přidává pomocí:

$$J_\lambda = J + \lambda \|\theta\|_2^2, \quad (19)$$

kde J reprezentuje vypočtenou celkovou ztrátovou funkci, θ je vektor parametrů a λ je regularizační koeficient (typicky velmi malý). Cílem je preferovat menší modely s menšími váhami. Dnešní implementace optimalizátorů zavádějí tuto operaci odděleně od výpočtu gradientu, jako je například u *AdamW* [14].

2.5.3.3 Dřívější zastavení

Při této technice se trénování ukončí dříve, za předpokladu, že se validační ztráta nelepší (po určitém počtu běhů – epoch). Využívá se tím faktu, že model generalizuje lépe předtím, než se dostane na minimální hodnotu ztrátové funkce. Zabráníme tím také přeučení a ušetříme čas běhu.

2.5.4 Stabilizační techniky

Během trénování může dojít k určitým numerickým nestabilitám. To jsou například vybuchující či mizející gradienty. Proto existují techniky, které se snaží dané problémy kontrolovat. Jedná

se například o *gradient clipping*. Při něm se při zpětném chodu učení kontroluje, zda norma gradientu není větší než daná prahová hodnota. Pokud k tomu dojde, dojde ke snížení její poměrné části (pomocí normy a daného prahu). Dalším takovým přístupem je využití normalizačních vrstev, které se také nachází v původním návrhu modelů Transformerů. Zlepšuje to nejen stabilitu, ale také urychluje konvergenci.

2.6 Varianty modelů

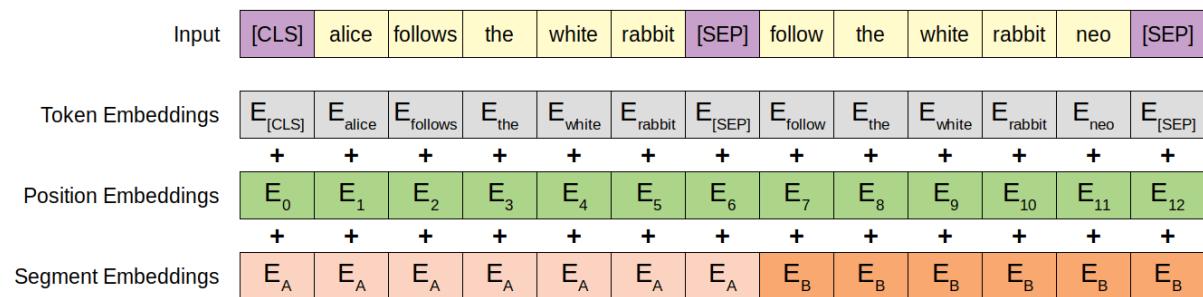
Málokterý model dnes využívá pouze základní strukturu. I dnešní nejpoužívanější modely (různé *LLM – Large Language Model*) využívají různá rozložení vrstev a cíle učení. V této části si představíme některé z nich.

2.6.1 BERT

Plným názvem *Bidirectional Encoder Representations from Transformers*, byl vyvinut společností Google v roce 2018. Představil obrovský skok v zpracování jazyka, kdy model se zvládl naučit hluboké porozumění kontextu slov ve větě. Dnes se bere jako výchozí srovnávací bod v experimentech *NLP*.

Jak už z celého názvu vyplývá, jedná se o architekturu využívající pouze enkodér. Struktura samotného enkodéru je ale stejná jako z práce Vaswaniho a kolegů. Počet vrstev se pak odvíjí od velikosti modelu (nejmenší začíná na 2 a největší končí na 24 vrstvách). Celkový model by se dal rozdělit do několika procesů: tokenizace, vnoření slov, vrstvy enkodéru a hlavy úlohy (*task heads*).

Tokenizace využívá *WordPiece*, ten rozdělí větu na menší jednotky slov či podmnožin slov (ne vždy jsou to samotná slova, dělí se i předpona a kořen slova například). Vnoření slov pak zohledňuje více částí. Nejdříve je vnoření tokenů, kde se rozdělí podle typu. Následuje poziční kódování, které transformuje sekvenci tokenů na matici čísel reprezentující pozici daného tokenu (původně slova). V modelu *BERT* se toto kódování zabaluje do procesu vnoření, protože je dynamicky měněno v rámci učení (oproti původnímu modelu, kde je pevně dané). Poslední úpravou je rozdělení podle segmentu. Vytvoří vektor podle pozice tokenu v rámci segmentu (jestli leží před nebo za speciálním tokenem, jedná se o binární reprezentaci). Celý proces je pak graficky znázorněn na obrázku 8. Takto připravená data pak přechází do enkodéru. Výstup z něj pak zpracuje hlava úloh na typy tokenů. Jedná se o opačnou operaci, k vnoření.



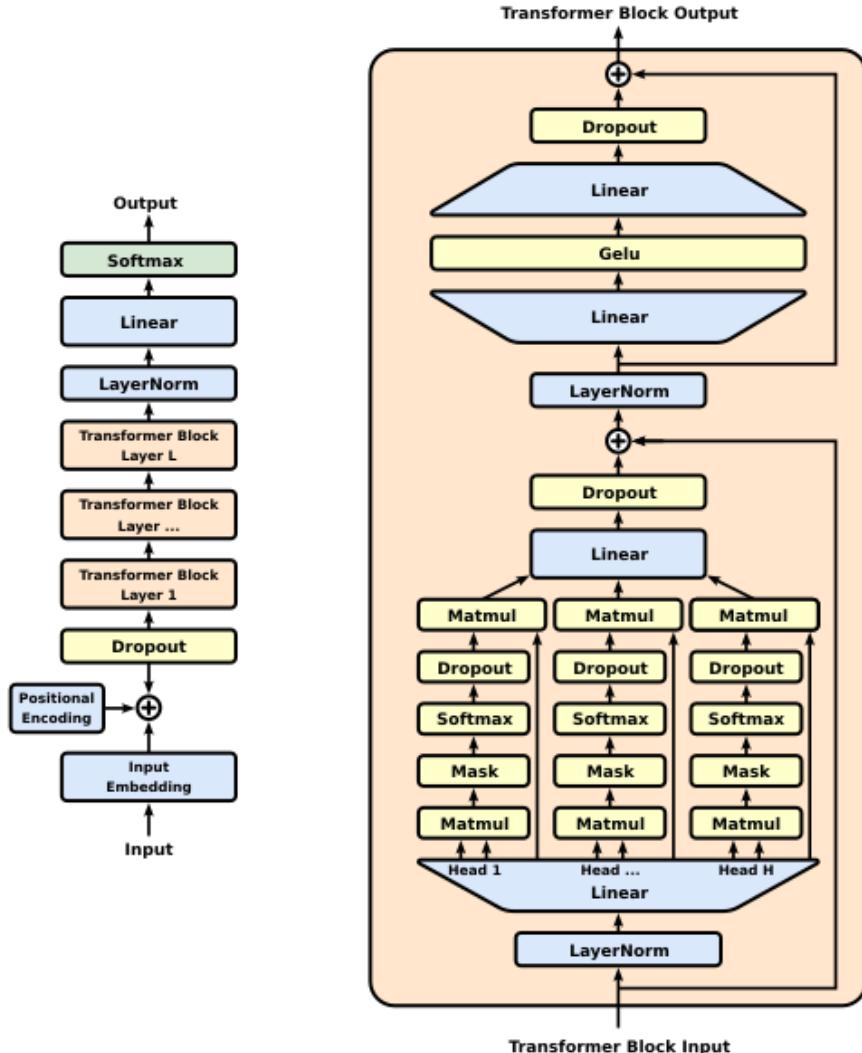
Obrázek 8: Nákres způsobu vnoření používaného v modelu BERT, převzato z [43]

BERT modely jsou koncipovány pro použití jejich předtrénovaných verzí a ty pak následně „dotrénovat“ na určité sadě dat pro daný problém (tzv. *fine-tuning*). Předtrénování se pak řídí strategií podle *MLM* (popsáno výše) nebo podle *NSP* (jedná se o proces, kdy se model snaží zjistit, zda dvě věty logicky navazují). *Fine-tuning* už probíhá na menším datasetu, se zaměřením na určitý problém (klasifikace textu, odpovědi na otázky apod.) [8].

2.6.2 GPT

Model GPT, celým názvem ***Generative Pre-trained Transformer***, byl vyvinut skupinou *OpenAI* v roce 2018 (verze GPT-1). Představila velmi schopný model v generaci textu, hlavně v ohledu produkce odpovědí na základě daného textu na vstupu. Dnes se stal klíčový modelem v oblasti generativního zpracování jazyka.

Architektura využívá pouze dekodér. Ten oproti původní skladbě neobsahuje druhou nemaskovanou vrstvu pozornosti (tam kde se připojuje výstup enkodéru). Tím, že využívá pouze maskovanou variantu, tak model podporuje autoregresivní (sekvenční) generování textu. Podrobnou strukturu lze vidět na obrázku 9.



Obrázek 9: Znázornění dekodér bloku používaného modely GPT, převzato z [44]

Tokenizace probíhá pomocí upravené verze *Byte-Pair Encoding*, který obdobně jako u *BERT* rozděluje větu na slova či části slov. Tokeny jsou však delší a nepřidávají indikaci, zda jsou tokeny původně jedno slovo. Rozdělení probíhá na bázi spojování nejfrekventovanější sousedních symbolů. Vnoření slov má pouze dvě fáze, přes tokeny a přes výpočet pozice. Nevyužívá segmentové rozdělení, protože se model netrénuje na *NSP*, není potřeba segmentová pozice. Výstup z poslední vrstvy je vektor, který je projekcí na slovníkovou velikost.

Obecně je pak model trénován na velkém datasetu za účelem predikce dalšího tokenu. Tím pádem není potřeba pro učení struktury jazyka ruční anotace dat. Stejně jako u BERT modelů se předtrénují na velkých datech a poté se pomocí *fine-tuning* upraví pro požadovanou úlohu [9].

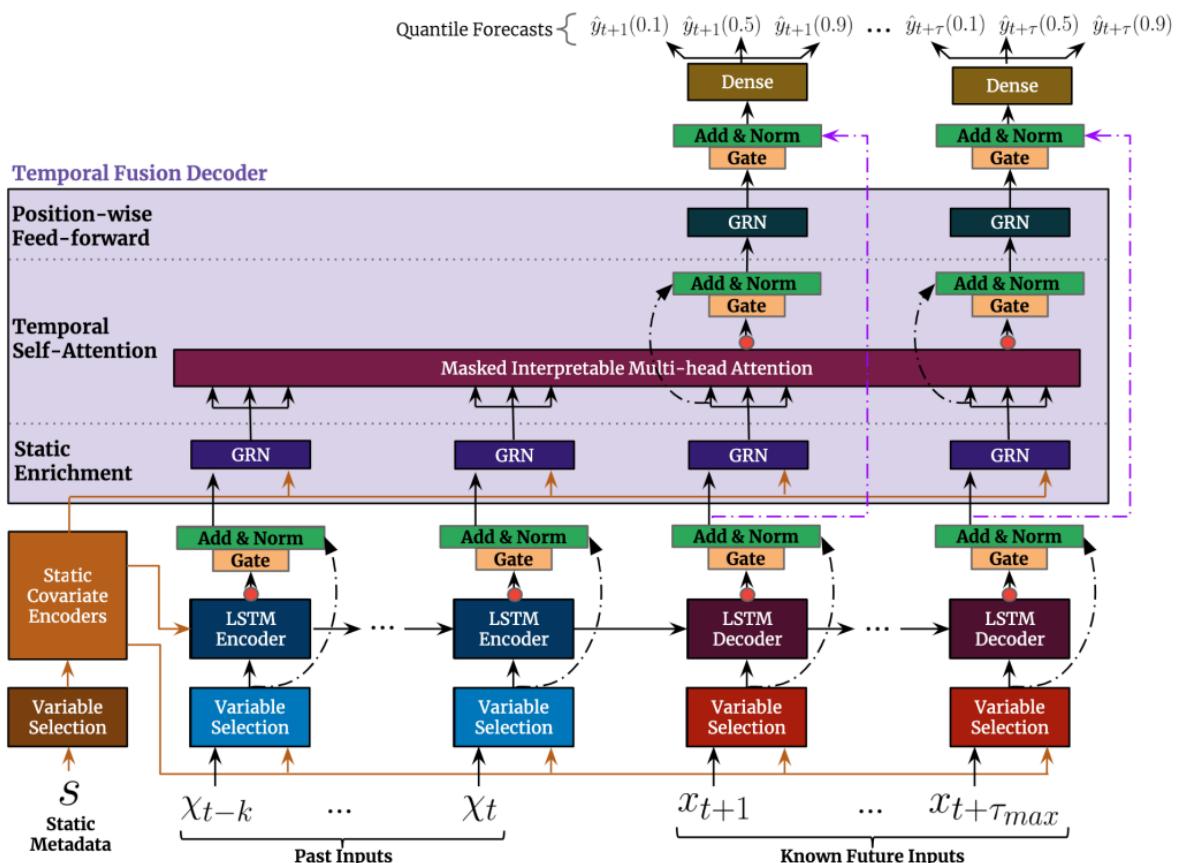
2.7 Využitý model

Pro predikci časových řad se využívá celá škála různých modelů. Nejčastějším a nejjednodušším zástupcem jsou různé enkodérové modely. Model použitý v této práci vychází z modelu *Temporal Fusion Transformer*.

2.7.1 Temporal Fusion Transformer

Tento přístup byl představen v práci z roku 2019 od autorů Bryan Lim, Sercan O. Arık, Nicolas Loef a Tomas Pfister [15]. Práce je zaměřena na predikce časových řad na delším horizontu, tedy předpovědi více časových kroků najednou. Zároveň se model snaží o co největší transparentnost, průběhu procesu při vytvoření výsledku (interpretovatelnost je důležitá vlastnost pro modely ve zdravotnictví nebo energetice). Predikce jsou tvořeny na základě dat ve složení:

- Minulých hodnot predikované proměnné
- Exogenní vlastnosti složené z neznámých i známých vstupu
- Statické kovariáty, určité neměnné charakteristiky pro různé časové řady



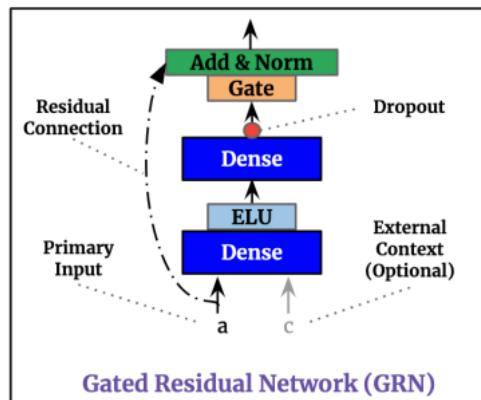
Obrázek 10:Nákres modelu TFT z původní práce, převzato z [15]

Grafické znázornění z původní práce pak lze vidět na obrázku 10. Ten zobrazuje jednotlivé části struktury modelu. Hlavními částmi jsou:

- Mechanismus s bránou (*Gating mechanism*)
 - Dodává modelu flexibilitu
 - Pomocí přeskakování (*skip*) nevyužitých komponent struktury
- Variabilní rozhodovací síť
 - Vybírá relevantní vstupy
- Enkodér pro statické kovariáty
 - Přidáním statické informace do sítě
- Časové zpracování (*temporal processing*)
 - Pro naučení dlouhodobých i krátkodobých časových vazeb
 - *Seq2seq* vrstva pro lokální informace a *multi-head attention* vrstva pro dlouhodobější vazby
- Výstup v delším horizontu
 - Kvantilní předpověď pro vytvoření požadované předpovědi v daném horizontu

2.7.1.1 Gated residual networks

Jedná se o implementaci mechanismu s bránou (*gating mechanism*). Vztah či vliv exogenních informací na predikovanou hodnotu nemusí být přesně známý a v některých případech může nepříznivě ovlivnit výsledek predikce. Úkolem tohoto prvku je zajistit, aby bylo nelineární zpracování aplikováno pouze na relevantní části vstupu – tedy na ty, které mají skutečný přínos pro danou predikční úlohu.

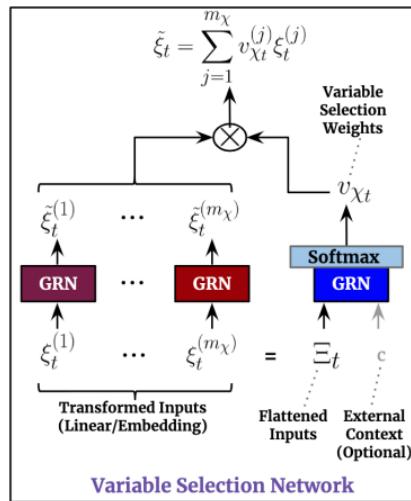


Obrázek 11: Struktura sítě GRN, převzato z [15]

Na obrázku 11 lze vidět strukturu *GRN* sítě. Architektura se skládá ze dvou plně propojených vrstev, mezi nimiž je použita *ELU* aktivační funkce (*Exponential Linear Unit*), která umožňuje i záporné výstupy, čímž pomáhá při řešení problému "mrtvých neuronů" a zlepšuje stabilitu učení. Druhá plně propojená vrstva má navíc aplikované odpadnutí (*dropout*). Následně informace pokračuje do vrstvy s bránou a skip mechanismem, který ovládá reziduální připojení. Brána se skládá z lineární vrstvy a *sigmoid* funkce, *skip* vrstva je lineární vrstva s podmírkou na základě dimenzí vstupu a výstupu (podle které se rozhoduje, zda má vliv na celkový výstup z brány). Na základě výstupu z brány a *skip* větve se spočítá výstup celé vrstvy s bránou, který rozhoduje, jak velká část informací se přenese dál a co se případně přeskočí.

2.7.1.2 Variabilní rozhodovací síť

Pro každý typ vstupních dat je tato síť vytvořena vlastní. Pro diskrétní (kategorické) hodnoty užívá *entity embedding* a pro spojité se jedná o lineární transformaci. Pomocí *softmax* funkce, GRN vrstvy a kontextu ze statických kovariátů (zahrnutí je nepovinné) se vypočítá vektor vah pro selekci proměnných. Zároveň s touto operací každá proměnná prochází vlastní GRN vrstvou. Tyto zpracované proměnné se pak spojí s vektorem vah pomocí sumy a vznikne agregovaný vektor vstupu pro daný časový krok. Touto aggregací dojde k filtraci proměnných. Popsanou strukturu lze vidět na obrázku 12.



Obrázek 12: Grafické znázornění operace selekce proměnných, převzato z [15]

2.7.1.3 Statické kovariáty

Statické informace se do modelu promítají skrze enkodéry pro ně navržené. Ty obsahují GRN vrstvy a výstupem jsou kontextové vektory pro další obohacení jiných kroků (rozhodovací síť, místní zpracování časových vlastností atp.).

2.7.1.4 Interpretable multi-head attention

U klasické varianty má každá hlava své váhy pro klíče, dotazy a hodnoty. Tím se komplikuje sledování poslední hodnoty. Proto TFT používá pro všechny hlavy stejné váhy hodnot. Díky tomu je jednodušší sledování, jak model pomocí mechanismu pozornosti vytvoří určitý výsledek.

2.7.1.5 Temporal fusion decoder

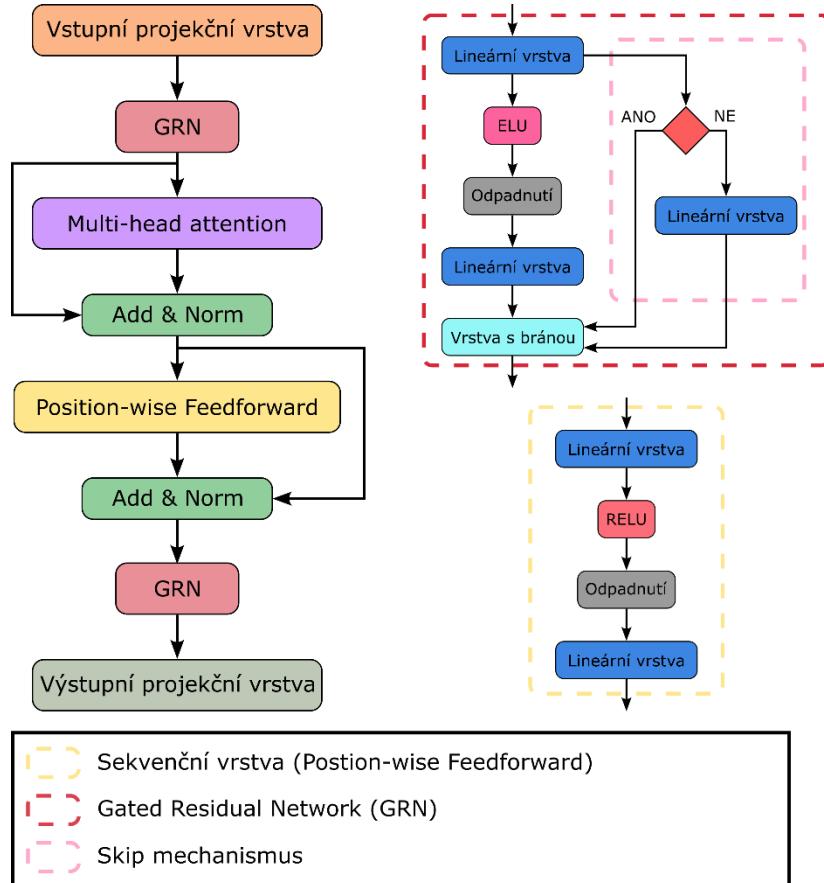
TFT model nahrazuje poziční kódování vrstvou seq2seq, která má lepší vlastnosti pro zachycení informace u časových řad. Struktura je obdobná jako u modelu využívající pouze dekodér, tedy chybí prostřední nemaskovaná vrstva s mechanismem pozornosti, na kterou se napojuje v klasické struktuře Enkodér. Zároveň však u plně propojených vrstev přidává mechanismus s bránou a s tím i GRN vrstvy.

2.7.1.6 Výstup v delším horizontu

Model pak výstup generuje jako distribuci hodnot vycházející z kvantilové ztrátové funkce (také známé jako *pinball loss*). Cílem je minimalizace přes všechny kvantilové ztráty všech výstupů [15].

2.7.2 Zjednodušená verze

Cílem modelů v této práci je predikce pouze jedné hodnoty, a proto některé části modelu byly odebrány. Zároveň se využívají předem známá data a není potřeba exogenní vstupy rozdělovat. Statické kovariáty byly také zanedbány. Celkovou strukturu modelu lze vidět graficky znázorněnou na obrázku 13.



Obrázek 13: Nákres zjednodušeného modelu TFT, červený kosočtverec pak představuje skip mechanismus

3 Současné využití ve finančním sektoru

Transformery prošly jako lavina obory zabývajícími se zpracováním přirozeného jazyka, od generování, upravování až po překlady či summarizace. Prakticky vytlačily všechny ostatní architektury a dnes se experimentuje hlavně s jejich modifikacemi. Tyto modely (a další) se pak využívají i v ekonomickém oboru. Rešerše se bude zabývat různými způsoby, jakými jsou používány Transformery ve finančním sektoru, speciálně se zaměřením na predikce časových řad.

3.1 Sentimentální analýza

Zpracování přirozeného jazyka (NLP) v oblasti financí je často zastoupeno v rámci analýzy sentimentu trhu. Mezi to patří i práce *Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex)* [16]. Ta kombinuje výhody lexikonových metod s Transformery. Přesněji používají *XLex* s transformery a metodou *SHAP* k automatickému rozšiřování finančních lexikonů. Tím pak snižují potřebu ruční anotace a zlepšují i přesnost klasifikace samotného sentimentu. Model založený na *XLex* (s využitím *RoBERTa*), společně s upravenými lexikony, dosahuje přesnosti 84,3 %.

Dalším zajímavým přístup je hybridní model *TRABSA* (*Transformer and Attention-based Bidirectional LSTM for Sentiment Analysis*) [17]. Ten kombinuje Transformer *RoBERTa*, mechanismus pozornosti a obousměrné *LSTM* sítě. Transformer jako takový má za úkol prvotní extrakci vlastností. Mechanismus pozornosti zvýrazní části textu obsahující nejvíce informací a obousměrný *LSTM* tyto upravené vlastnosti zpracuje. Tím se dosahuje lepšího zachycení časových spojení a samozřejmě i přesnější klasifikace. Přesnost tohoto modelu pak činí 94 %.

Neexperimentuje se pouze s modely na bázi architektury BERT, například práce *FinLlama: Financial Sentiment Classification for Algorithmic Trading Applications* [18] využívá model *Llama*. Autoři upravili model *Llama 2 7B* na použití pro finanční data. Přístup pak využívá techniku *LoRA* v rámci *fine-tuningu* modelu (tím snižuje výpočetní náročnost bez ztráty přesnosti). Tento model pak na základě analýzy sentimentu vytvořil portfolio a jeho výsledky porovnal s dalšími využívanými modely, například model *FinBERT*. Roční návratnost modelu pak činila 45 % (o 15 % více než *FinBERT*).

3.2 Detekce podvodů a anomálií

Dalším zajímavým užitím Transformerů je detekce podvodů a anomálií. Práce *From anomaly detection to classification with graph attention and transformer for multivariate time series* [19] se zabývá touto problematikou. Využívá metodu grafového učení (*GNN*) a Transformeru (*Informer*). Nejdříve je Transformer použit na rekonstrukci vstupu rozděleného na časová okna. Oba zmíněné modely pak pokračují do adverzálního učení, které je podobné *GAN* modelům. Model pak v nejlepší konfiguraci řeší problém s 85% přesností.

Další práce navrhuje metodu na základě architektury Transformerů nazvanou *Variable Temporal Transformer (VTT)*. Využívá mechanismus *self-attention* ke zpracování časových závislostí a vztahů mezi daty. Model kombinuje časovou pozornost a variabilní pozornost pro modelování korelací mezi senzory. Zároveň je k modelu přidán modul pro interpretaci anomálií pro lepší čitelnost výstupu. Přesnost pak odpovídá 87 % [20].

3.3 Modelování časových řad

Modely, které správně předpoví další číslo v řadě (nebo více) jsou neocenitelným pomocníkem při obchodování na finančních trzích. Pro tvorbu takových modelů se využívají všechny různé metody strojového učení, a tedy i Transformerů.

Autoři práce *Timer: Generative Pre-trained Transformers Are Large Time Series Models* [21] navrhují upravený model, který nazvali *Timer (Time Series Transformer)*. V něm představili novou kategorii inspirovanou velkými jazykovými modely, *large time series models*. Ty by měly plnit obdobnou úlohu jako *LLM* aby i datasety založené na časových řadách neměly modely stavěné od základu (mohlo by se jedna o modely jako *BERT* a *GPT* pro časové řady a další experimenty by na nich mohly stavět). Tato sféra modelů má za cíl, stejně jako *LLM*, řešit více druhů problémů, což dělá i *Timer*. Ten je předtrénován na generaci další hodnoty, ale je následně adaptován i pro jiné úlohy, jako je imputace dat (doplňení chybějících dat) nebo detekce anomalií.

Dalším zajímavým modelem je *TimeXer*. Ten do učícího procesu zahrnuje exogenní proměnné mimo proměnnou sledovanou. *TimeXer* pak využívá strukturu založenou na dekodéru s klasickou skladbou vrstev. Exogenní data pak vstupují v rámci *cross-attention* v druhé vrstvě (stejně jako by vstupoval výstup enkodéru) a endogenní proměnná je na vstupu dekodéru [22].

Další prací zaměřující se na obecný model pro časové řady je *Unified Training of Universal Time Series Forecasting Transformers* [23]. Představuje koncept univerzální predikce časových řad *MOIRAI (Masked EncOder-based UnIveRsAl TIme Series Forecasting Transformer)*. Byl trénován na datasetu *LOTSA* s více než 27 miliardami pozorování z různých odvětví. Jedná se o enkodér architekturu s upravenými normalizačními vrstvami a plně propojenými vrstvami. Nejmenší verze měla naměřenou agregovanou normalizovanou absolutní odchylku rovnou 0,655.

Odlišnou cestou pak jde model *PATHFORMER*. Zavádí *multi-scale* architekturu s adaptivními cestami. Model rozděluje časové řady na více měřítek, čímž vytváří odlišná časová rozlišení. Využívá duální pozornost, která sleduje globální korelace a lokální závislosti. Adaptivní cesty pak upravují zpracování podle časové variability dat [24].

3.4 Předpovědi časových řad

Modely se také specifikují přímo na predikce bez použití pouze časových dat. Speciálně se tato část zaměří na předpovědi na finančním trhu. Práce *Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction* [25] využívá *Multi-Scale Gaussian Prior* (přidání Gausovské masky) pro zachycení lokálních vzorů. Dále ortogonální regularizace předchází vzniku redundantních hlav u mechanismu pozornosti. Využívá tzv. *Trading Gap Splitter* (dvě poziční masky pro rozdelení denních a týdenních časových kroků) pro extrahování rysů z finančních dat.

Dále máme hybridní model *TEANet*, který kombinuje Transformer a *LSTM*. Transformer je na bázi enkodéru a získává klíčové informace ovlivňující trh, které jsou zpracovány a zkombinovány s dalšími daty. Následně jsou předány *LSTM* s časovým pozornostním mechanismem. Využívá data ze stránky *Twitter* a ceny z trhů. Přesnost se pohybovala okolo 65 % [26].

Práce *Stock market index prediction using transformer neural network models and frequency decomposition* [27] využívá Transformery s konvolučními neuronovými sítěmi a frekvenční

dekompozicí *EMD* a *CEEMDAN*. Autoři ukazují, že rozklad časové řady na více frekvenčních složek s nižší volatilitou usnadňuje učení a zvyšuje přesnost.

Dalším modelem využívajícím dekompozice je *SDTP* (*Series Decomposition Transformer with Period-correlation*). Ten se soustředí na zpracovávání dat z trhu. Časovou řadu dělí na trendovou a sezónní složku. V rámci mechanismu periodické korelace seskupuje podobné časové úseky a aktualizuje informace agregací relevantních částí [28].

Nový hybridní model navrhla i práce *A novel hybrid model for stock price forecasting integrating Encoder Forest and Informer* [29]. Kombinuje přístup dekompozice a rekonstrukce společně s Transformer modely. Nejdříve data rozdělí na nízkofrekvenční a vysokofrekvenční složku a predikci zajišťují *Informer* a *Encoder Forest*.

Práce *A Stock Price Prediction Method Based on BiLSTM and Improved Transformer* [30] využívá taký hybridní model. Autoři využívají kombinaci obousměrné varianty *LSTM*, víceúrovňového transformeru a časových konvolučních sítí. Poslední jmenovaná nahrazuje dekodér, ve smyslu struktury Transformeru. Zaměření predikcí bylo na firmy z různých odvětví a indexů. Konkrétně se jednalo o 19 různých aktiv z čínského finančního trhu. Model v porovnání s konvenčními přístupy jako jsou *LSTM* projevil vylepšení ve střední kvadratické odchylce o 24,3 % a v koeficientu determinace o 15,6 %.

Spojení modelu *GRU* a Transformeru používají autoři v práci *Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators* [31]. Ta se zabývala predikcemi na finančním trhu použitím dat z novinových článků (indonéských) a různých technických indikátorů. Předpovídánými aktivity byly vybrané indonéské firmy s vysokou kapitalizací. Navrženou architekturu následně porovnaly pomocí metrik chyb s některými transformerovými modely (základní Transformer, Informer atp.).

4 Data

Správně zpracovaná data jsou klíčová pro odhalení informací v nich obsažených. Obecně pak platí, že s množstvím relevantních dat se i zlepšuje kvalita modelů. Data použitá pro predikci, v případě modelů této práce, se dají rozdělit do 3 kategorií: finanční trhy, technické a jiné zpožděné indikátory a investorova pozornost. Modely v této práci prozkoumávají predikce na více druzích aktiv a pro každou z nich je tvořen i dataset. Celkový časový rozsah dat je 1.1.2019 až 31.12.2024.

4.1 Finanční trhy

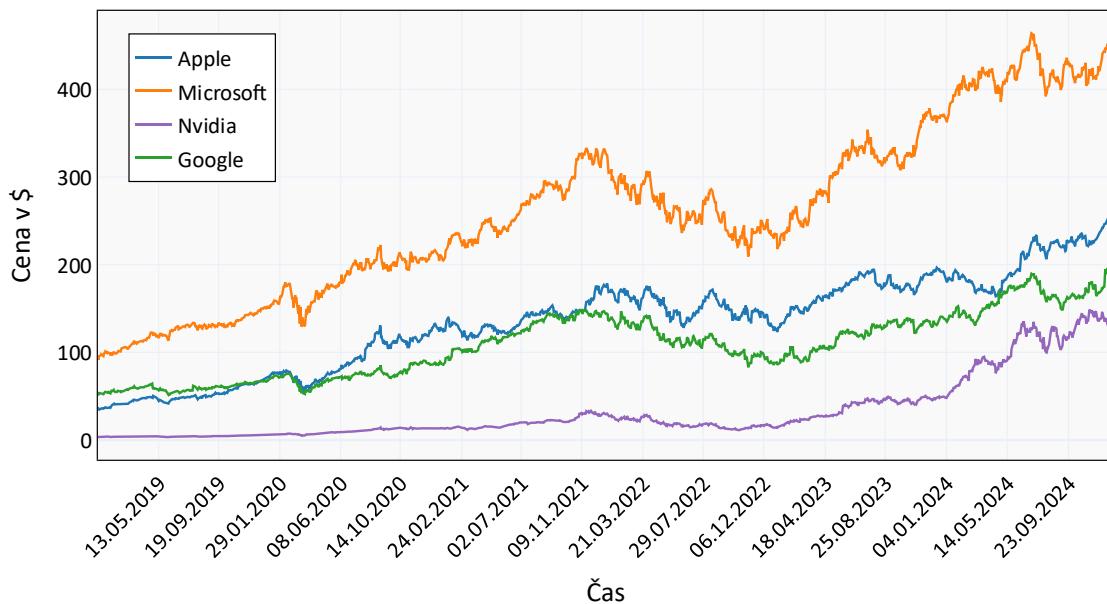
Kromě cen predikované komodity obsahují datasety také informace o dalších aktivech, jako jsou akcie, opce či měnové kurzy. Tato aktiva mohou souviset s cílovou hodnotou buď prostřednictvím reálných ekonomických vazeb (například konkurenční společnosti, komodity se stejným využitím nebo vzájemně ovlivňované měny), nebo prostřednictvím příslušnosti ke stejnemu investičnímu či ekonomickému sektoru (např. významná firma obchodovaná na burze, opce na hojně využívanou komoditu, případně kryptoměna používaná ve spekulativních strategiích). Datasety dále obsahují kurzy měn, které mohou ovlivňovat vývoj sledovaných hodnot, zejména v případě komodit obchodovaných v konkrétních měnách. Přesné složení datasetů a způsob výběru těchto atributů je podrobně popsán v následujících podkapitolách.

4.1.1 Apple

Nadnárodní korporace se sídlem v USA. Založena v roce 1976 Stevem Jobsem a Stevem Wozniakem. Začali jako neznámá společnost s ambiciózní vizí v technologickém sektoru. V témže roce začali i s výrobou prvního počítače (nejdříve pouze základní deska s CPU a RAM). O rok později pak přišel druhý, již o dost známější produkt, Apple II, který byl i komerčně velmi úspěšný. Tím největším průlomem byly ale Apple Lisa (technologicky) a Apple Macintosh (technologicky i komerčně), které přidaly nové uživatelské rozhraní používající myš (nápad firmy Xerox, která se s firmou Apple i soudila, avšak neúspěšně). Díky tomu se z nich stala firma se světovým renomé. Firma pak procházela různými změnami, od odchodu Jobse po různé aliance s ostatními technologickými firmami (IBM, Motorola). V tomto období si pak firma nevedla tak skvěle. Po různých peripetiích (změny vedení, nákup nové firmy Jobse NeXT) se opět hlavním ředitelem stal Steve Jobs. S tím přišla i revitalizace firmy a identity značky jako takové. Začala produktová řada iMac, v roce 2001 iTunes spustil revoluci v hudebním průmyslu (společně pak s prodejem iPodů) a hlavní pak byl v roce 2007 příchod telefonů iPhone s dotykovým displejem. S tímto zařízením pak v rámci boje s konkurencí (hlavně BlackBerry) přišli s celou škálou různých funkcionalit, které se staly průmyslovým standardem. V roce 2010 pak představili iPad jako kombinaci laptopu a chytrého telefonu. Dnes firma vyniká nejenom v produkci různých zařízení (telefony, počítače, hodinky) či jejich příslušenství (periférie, doplňky atp.), ale i v softwarové stránce (internetová televize, trh s aplikacemi nebo hudbou, cloudové úložiště). Zároveň se jedná o jednu z firem s největší kapitalizací na světě (místa se mění, ale pohybuje se v rámci prvních tří míst, společně s Nvidií a Microsoftem) [32].

V tomto případě se dataset zaměřuje na predikci konečné (Close) ceny akcií Apple. Kromě pohybů cen akcií této společnosti pak obsahuje tyto informace o firmách Microsoft, Nvidia a Alphabet (dříve společnost Google). Ty zastupují různé konkurenty v technologických odvětvích. Google či Microsoft jsou dnes spíše softwarově zaměřenými firmami (Microsoft v minulosti více stavěl i na hardware, dnes již není hlavním produktem, i když existují věci jako řady počítačů Surface nebo Xbox herní konzole), každá má dominantní postavení v jiných oblastech.

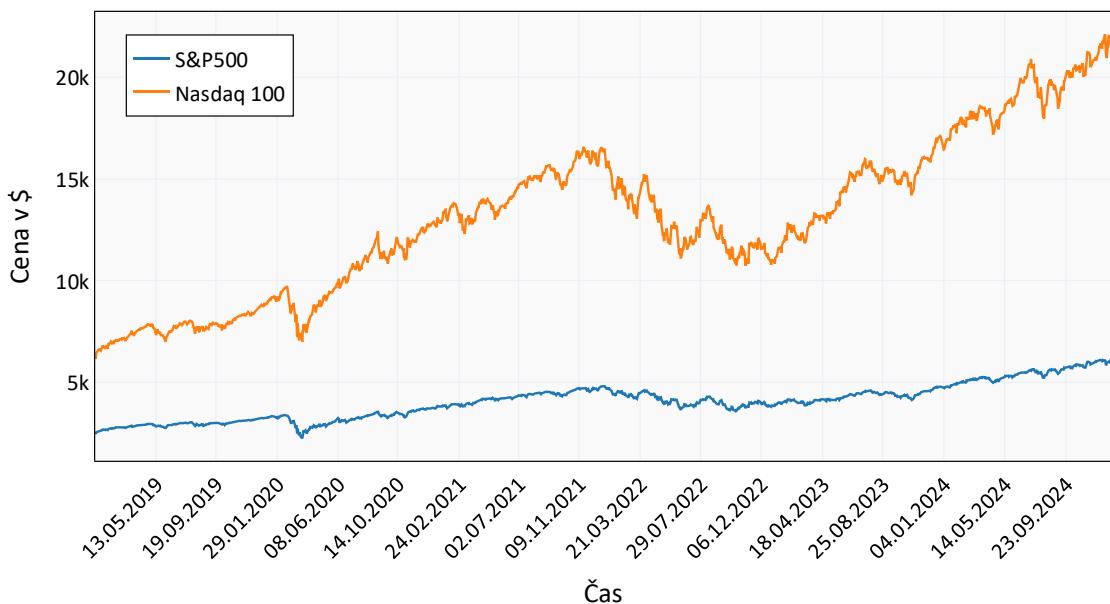
Ceny zahrnutých akcií



Obrázek 14: Graf vývoje cen akcií velkých firem přes celé sledované období

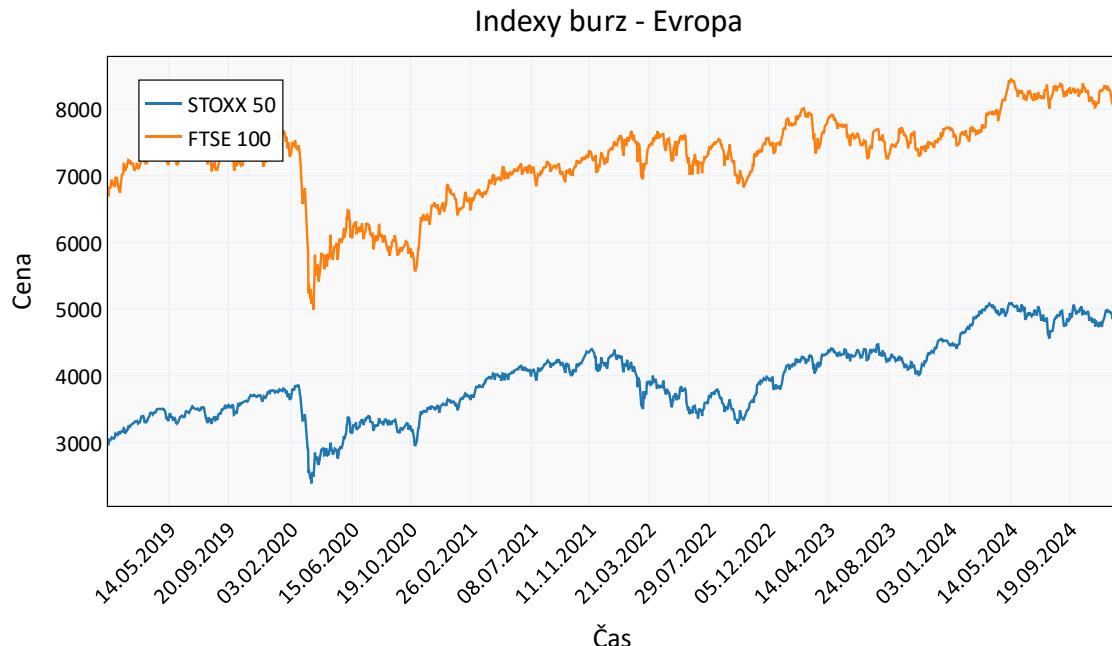
Zároveň obsahuje i zastoupení různých indexů finančních trhů. *S&P 500* a *Nasdaq 100* jsou nejsilnějšími zástupci se sídly v New Yorku. Obsahují nejsilnější firmy nejen v USA, ale i na světě. *Nasdaq* se zaměřuje hlavně na IT, biotechnologický a papírenský sektor. 50 největších firem z indexu *S&P 500* pak obsahuje více než polovinu kapitalizace celého trhu. Vývoj pak lze vidět na obrázku 15.

Indexy burz - USA



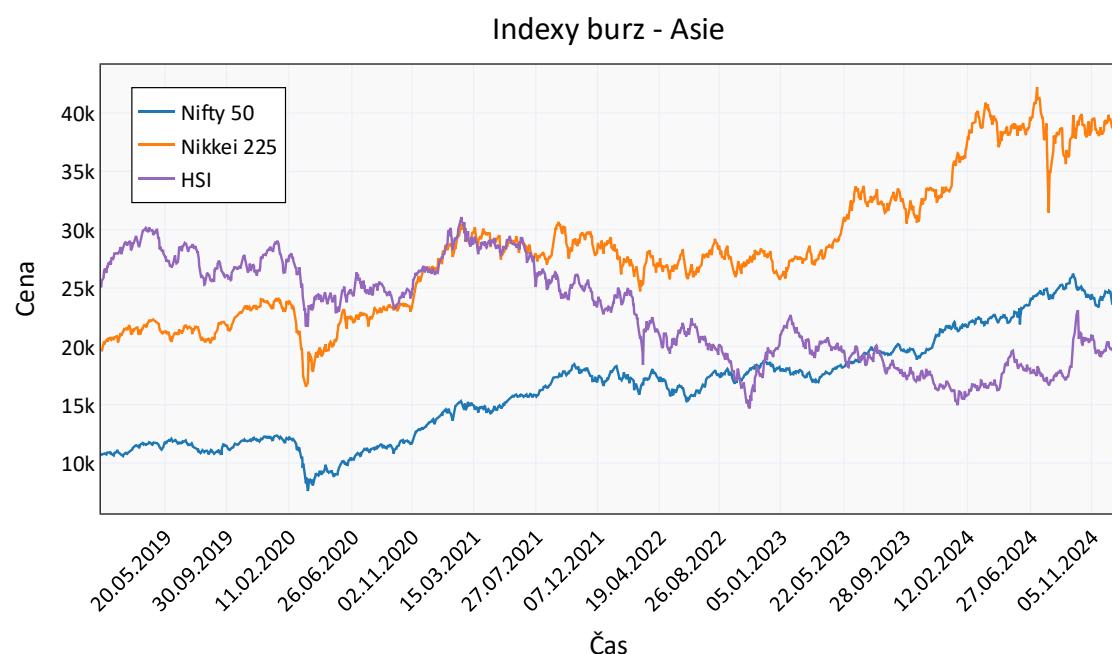
Obrázek 15: Graf indexů s největší kapitalizací na světě, oba s centry v USA

Evropu pak zastupují indexy *FTSE 100* a *STOXX 50*. První jmenovaná obsahuje 100 největších firem britského trhu s cennými papíry v Londýně. Druhý spojuje 50 firem eurozóny se zaměřením na „*blue chip*“ společnosti (spolehlivé, silné, lídry vlastních sektorů). Graf ceny je vyobrazen na obrázku 16.



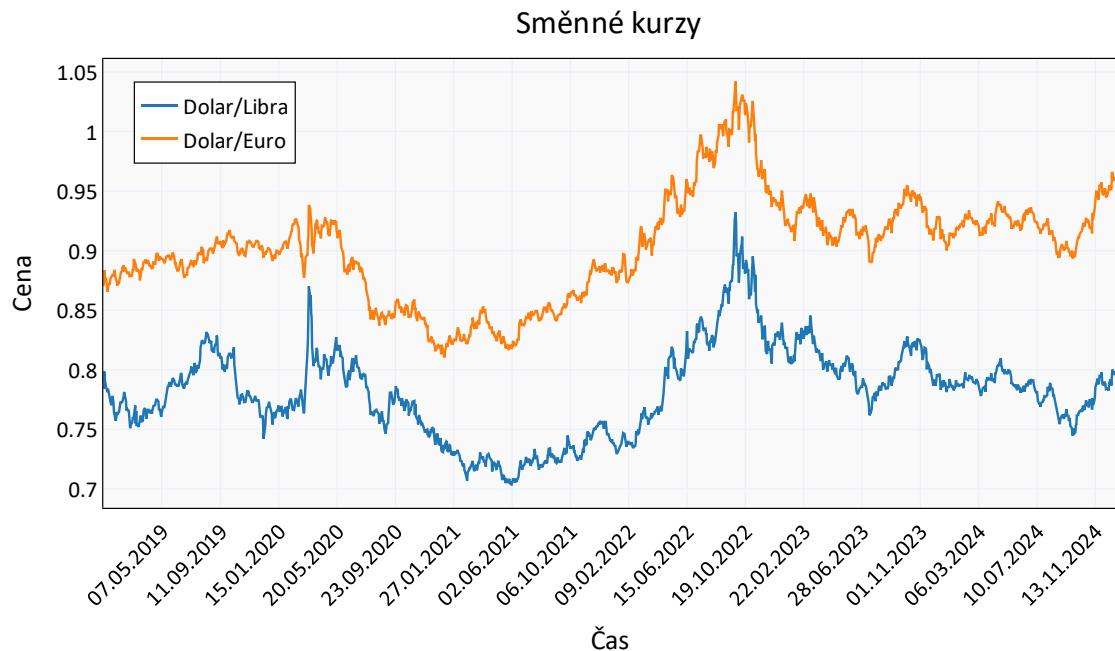
Obrázek 16: Hladiny cen indexů pocházejících z Evropy

Asii reprezentují 3 různé indexy. *Nifty 50* je index indické burzy obsahující 50 největších firem přes 13 sektorů (největším jsou banky a finanční instituce). *Nikkei 225* je japonský burzovní index se sídlem v Tokiu. Obsahuje 225 veřejně obchodovaných firem. *HSI* je index hongkongské burzy a obsahuje 82 firem. Porovnání jejich cen lze spatřit na obrázku 17.



Obrázek 17: Graf indexů nejsilnějších burz z Asie

Posledními v datasetu jsou kurzy různých měn a ceny opcí zlata. Přesněji se jedná o kurz amerického dolara k euru a amerického dolara ku britské libře. Porovnání vývoje těchto kurzů v čase lze vidět na obrázku 18.

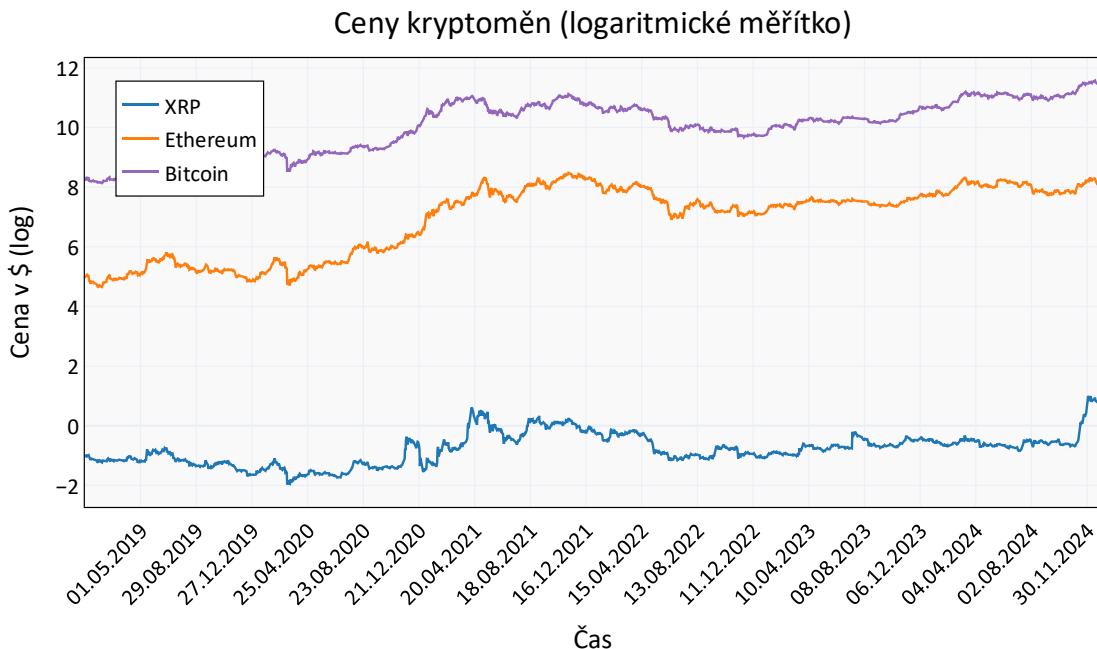


Obrázek 18: Směnné kurzy v období 2019-2024

4.1.2 Bitcoin

Bitcoin patří mezi nejstarší kryptoměny. Vznikl v roce 2009 (popsán byl již o rok dříve) a autorem je uveden Satoši Nakamoto (není známo, zda se jedná o skupinu či jednoho autora). Jeho hlavní výsadou (a většiny kryptoměn) je decentralizace, aby ani autor nemohl ovlivňovat toky či vyvolat inflaci. Z názvu této vlastnosti vyplývá, že měna nemá žádný centrální bod, transakce se dělají pomocí peer-to-peer distribuované databáze. Díky tomuto systému je přesně daný celkový počet měny a nelze ho navýšit (jako je to obvyklé u normálních měn, kdy centrální banky mají změnu peněžní zásoby jako součást své monetární politiky). Nové Bitcoiny se pak tvoří tzv. těžením, kde skupiny „závodí“, kdo první rozluští kryptografický problém spojený s procesem validace transakcí. První pak získá odměnu za toto úsilí a systém pokračuje. Každé 4 roky (zhruba) pak dochází k tzv. „halvingu“, to je krácení této odměny o 50 %. Dnes už některé obchody (především internetové) umožňují platit napřímo pomocí kryptoměn, avšak stále se ani zdaleka nejedná o alternativu ke klasickým měnám. Proto existují různé burzy pro převedení mezi nimi. Je spojován s nechvalně proslulými černými internetovými trhy z důvodu anonymity (samý o sobě lze transakce dohledat v blockchainu – databázi, existují však různé způsoby komplikování tohoto procesu, např. mixery). Dnes je díky své rychle rostoucí ceně a volatilitě brán jako zajímavý investiční artikl i pro konzervativnější investory a fondy.

Opět obsahuje firmy, opce ropy či zlata a kurzy dolaru k libře a euro (s přidáním kurzu jüanu). Novými částmi jsou vývoje jiných kryptoměn. Přesněji jde o cenu/kurz Ethereum a XRP (samozřejmě i ceny Bitcoinu samotného). Graf jejich cen přes celé sledované zobrazuje graf na obrázku 19 (v logaritmickém měřítku z důvodu velkých cenových rozdílů)



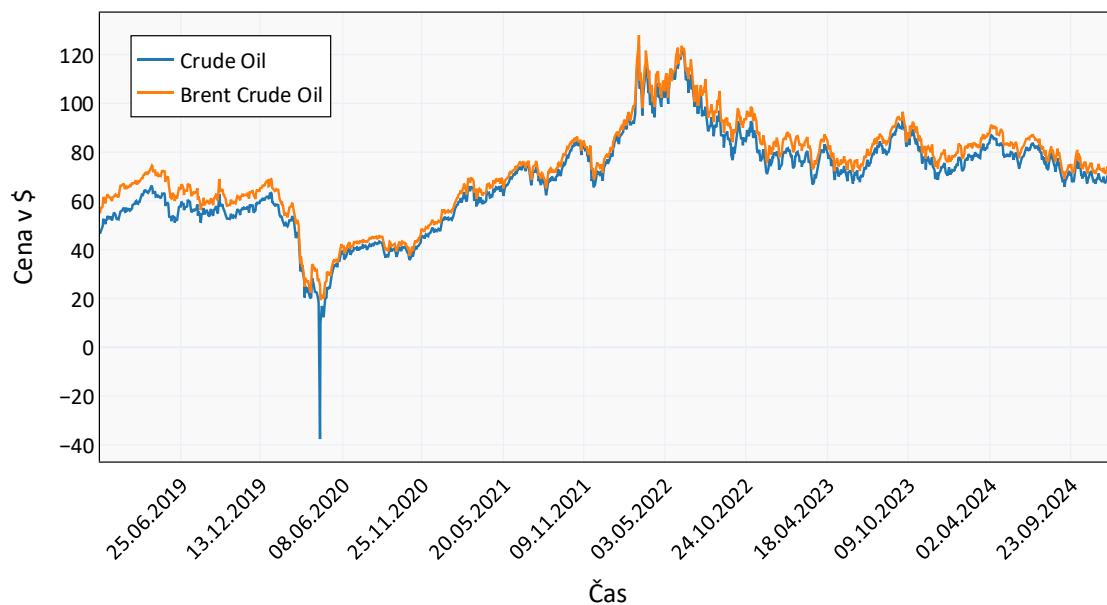
Obrázek 19: Ceny/kurzy kryptoměn v logaritmickém měřítku

4.1.3 Zlato

Zlato je jednou z nejstarších cenin. Dříve se využívalo jako měna nebo jako zlatý standard (vyjádření hodnoty v měnovém systému). Standard se dnes již nepoužívá, ale má i mezi ekonomickými proudy své podporovatele. Rakouská škola (ekonomická díla vídeňských autorů, klasický liberalismus a libertarianismus) je jednou z nejprominentnějších, protože má zabráňovat nadměrné tvorbě peněz a tím i inflaci. Využívá se v celé řadě elektronických zařízení, nejčastěji v konektorech a pinech. Z investičního hlediska je často viděno jako stabilní komodita, proto je často více nakupováno v časech recese. Může se tedy brát jako určitý ukazatel stavu trhu (může se však jednat i o jiné faktory).

Dataset opět obsahuje ceny firem Apple, Nvidia, Microsoft a Alphabet. Stejně tak kurzy měn či indexy světových burz. Rozdílné je, že obsahuje opce pro *Crude Oil* a *Brent Crude Oil*. I když oba označují futures opce ropy, jedná se o její jiné druhy. První jmenovaná zaštiťuje výsledek těžby na ropných polích v USA, druhá zahrnuje ropu z těžby převážně v Severním moři (Norsko, Dánsko a další). Jejich ceny jsou blízce provázány, jak je možno pozorovat na obrázku 20.

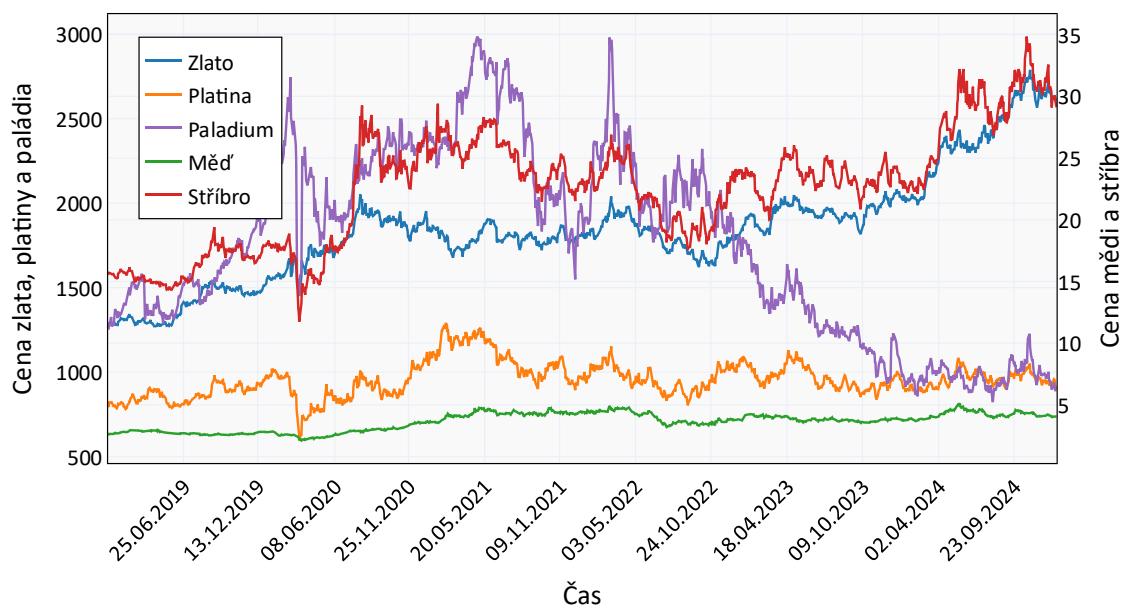
Ceny opcí ropy



Obrázek 20: Vývoj ceny hlavních dvou ropných opcí

Ceny ostatních cenných (či jiných) kovů jsou také důležité, hlavně z investičního hlediska. Proto data zohledňují ceny paládia, stříbra, mědi a platiny. Jejich porovnání lze vidět v oboustranném grafu na obrázku 21.

Ceny různých kovů



Obrázek 21: Ceny jednotlivých zahrnutých kovů, graf je rozdělen podle cen, kde zlato, platina a paládium patří k levé a měď se stříbrem k pravé

4.2 Technické indikátory a zpožděné proměnné

Využívají se pro lepší zachycení různých trendů a sezónností, zlepšují generalizaci a snižují zpoždění předpovědi. Dohromady přidávají 14 různých vlastností do složitosti modelu.

4.2.1 Klouzavé průměry

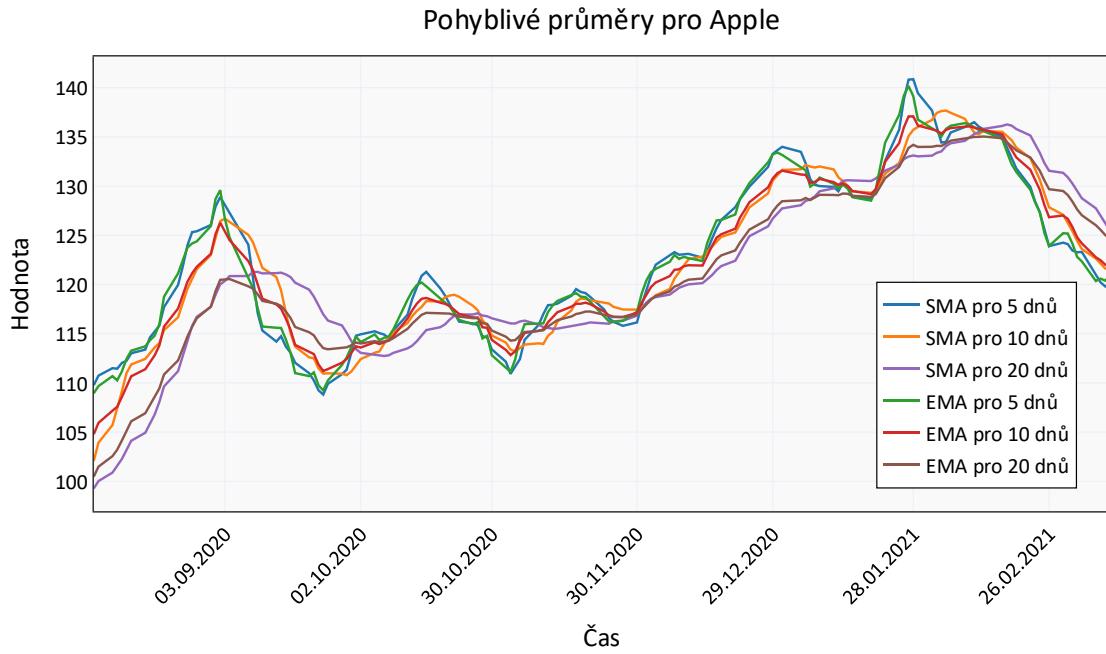
Jedná se o nejpočetnější skupinu, co do počtu vlastností. Dohromady jich tvoří 6. Počítají se pro 5, 10 a 20 dnů. Výpočet je prováděn z konečné ceny dané predikované komodity. První je jednoduchý klouzavý průměr:

$$SMA = \frac{A_1 + \dots + A_n}{n}, \quad (20)$$

kde n je velikost úseku řady a A_1 až A_n jsou prvky časové řady daného úseku. Dalším je pak exponenciální klouzavý průměr:

$$EMA_t = \left(A_t \cdot \left(\frac{s}{1+n} \right) \right) + EMA_{t-1} \cdot \left(1 - \left(\frac{s}{1+n} \right) \right), \quad (21)$$

kde A_t je prvek řady v kroku t , s je hodnota vyhlazení (typicky 2) a ne je opět velikost úseku [33]. Porovnání jejich vývoje lze vidět na obrázku 22.



Obrázek 22: Graf pojednatelných průměrů na ceně akcie Apple, zobrazuje kratší časový úsek pro lepší čitelnost

4.2.2 Bollingerova pásmá

Jedná se o metodu technické analýzy, která se využívá na identifikaci potenciálního přehnaného nákupu či prodeje a celkové volatility. Obsahuje tři komponenty: horní, střední a spodní pásmo. Pokud cena vystoupá či klesne nad/pod jedno z krajních pásem, může to poukazovat na zmíněný větší nákup nebo prodej. Zároveň se využívá jako jeden z indikátorů trendu (cena mezi středním a horním znamená potenciální rostoucí trend a naopak). Výpočet je definován:

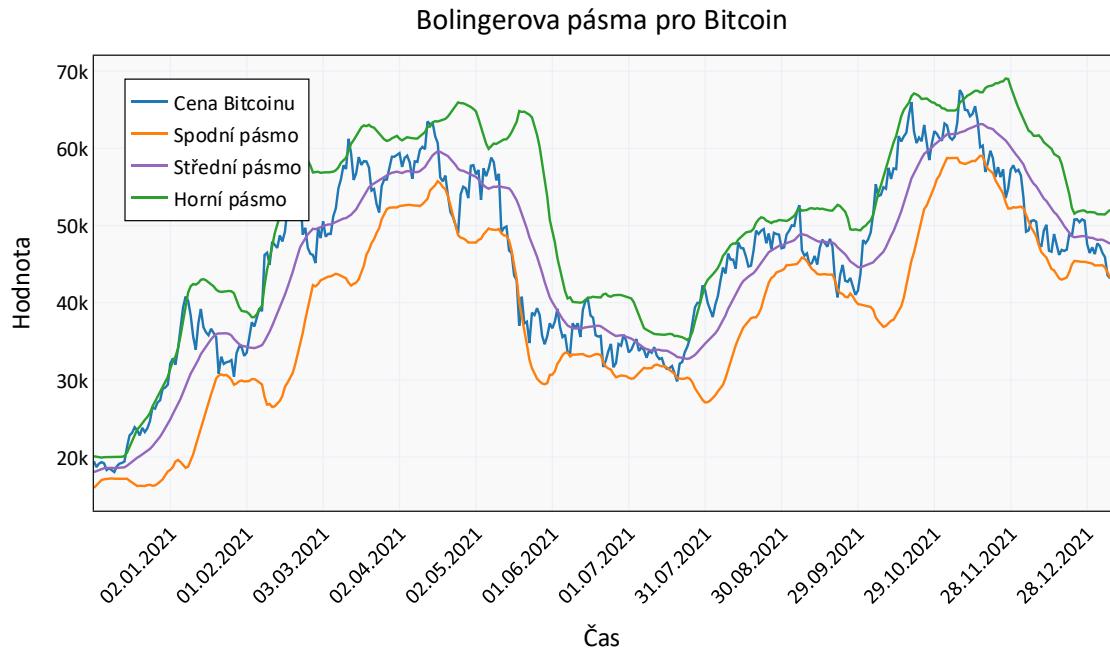
$$UpperBB = SMA_n + r \cdot \sigma_n, \quad (22)$$

$$MiddleBB = SMA_n, \quad (23)$$

$$LowerBB = SMA_n - r \cdot \sigma_n, \quad (24)$$

kde SMA_n je jednoduchý klouzavý průměr, σ_n je standartní odchylka, r je koeficient a index n označuje přes jaké období se výpočet jednotlivých částí provádí. Typicky r má hodnotu 2 a n je období 20 dnů [34].

Na obrázku 23 lze vidět Bollingerova pásmá pro apple pro rok 2021. Na něm lze vyzorovat vztah mezi pohybem ceny nad či pod středním pásem. Velice zřetelné je to na začátku května, kdy se cena dlouho pohybuje ve spodní části pásm. Během dalšího měsíce vykazuje silný klesající trend.



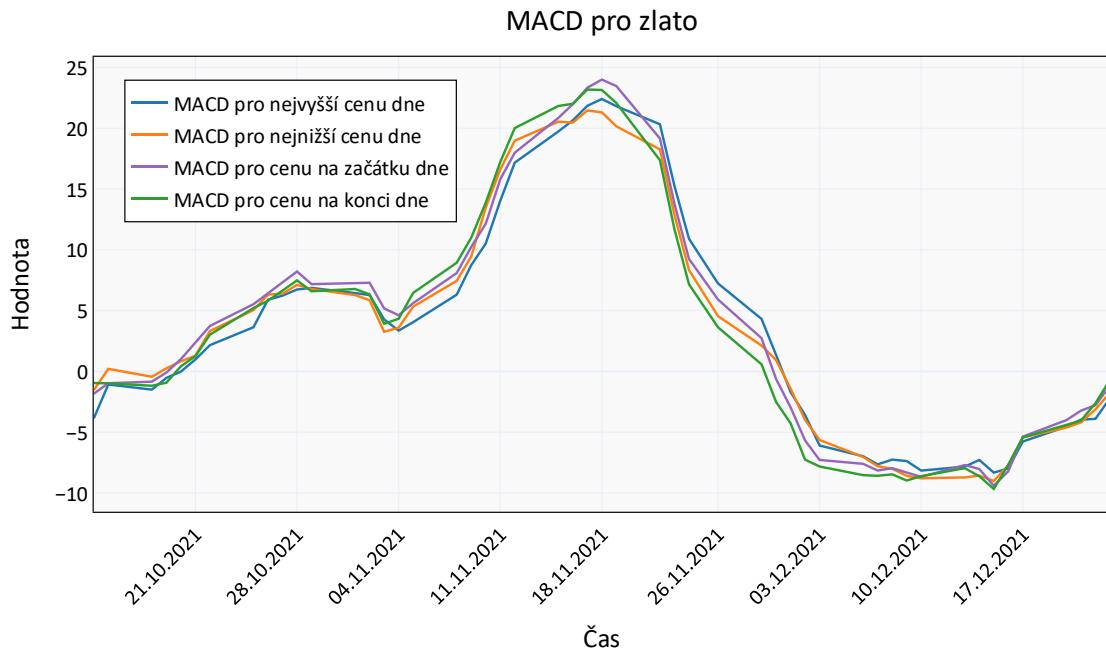
Obrázek 23: Bollingerova pásmá společně s cenovou hladinou akcie Apple

4.2.3 Moving Average Convergence Divergence (MACD)

Představuje další indikátor technické analýzy. Obecně se skládá ze tří komponent: MACD linky, signál linky a histogramu. Dataset jako takový obsahuje MACD linky pro jednotlivé složky ceny akcie (myšleno Low, Open, High a Close). MACD linka je pak vypočtena jako:

$$MACD \text{ line} = EMA_{12} - EMA_{26}, \quad (25)$$

kde EMA označuje exponenciální klouzavý průměr (12 a 26 dnů podle indexu). Signální linku lze vypočítat jako exponenciální klouzavý průměr z $MACD \text{ line}$. Tento technický indikátor se pak využívá pro odhad celkových trendů. Při manuální technické analýze se sledují body průniku signální a MACD linky a na základě histogramu lze pak odhadovat, zda se bude jednat o medvědí (klesající) nebo býcí (stoupající) trh [35].



Obrázek 24: $MACD \text{ line}$ pro všechny denní ceny zlata na trhu, zobrazuje pouze úsek dat pro lepší čitelnost

4.2.4 Index Relativní síly

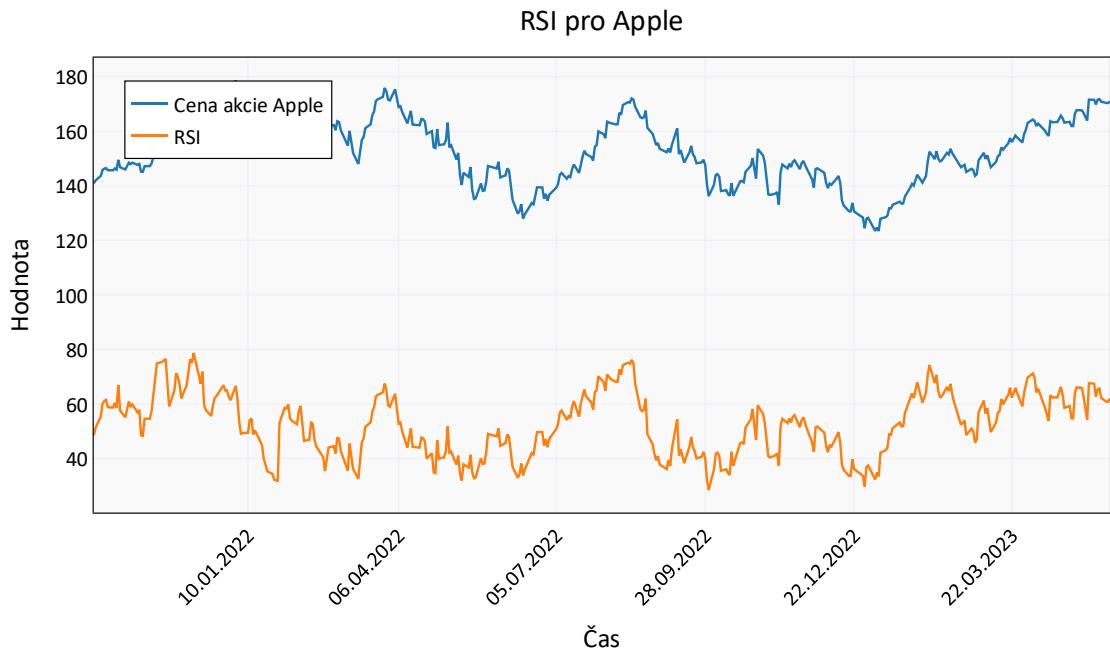
Jedná se o technický indikátor se zaměřením na výpočet rychlosti a velikosti změny ceny a předpověď přehnaného nákupu nebo prodeje. Nejdříve se vypočítá relativní síla:

$$RS = \frac{EMA_n(U)}{EMA_n(D)}, \quad (26)$$

kde U označuje pozitivní změnu v ceně (rozdíl dnešní ceny a včerejší; pokud je rozdíl záporný, vrátí funkce 0), D je negativní změna v ceně (obdobně jako u pozitivního rozdílu), EMA_n označuje exponenciální klouzavý průměr přes n dnů. Index samotný lze získat jejím využitím [36]:

$$RSI = 100 - 100 \frac{1}{1 + RS}. \quad (27)$$

Na obrázku 25 lze vidět graf RSI společně s cenou Apple.



Obrázek 25: Graf RSI s cenou akcie Apple přes celý dataset

4.2.5 Fourierovy transformace

Posledním druhem zpožděných proměnných jsou Fourierovy transformace. Přidávají další vrstvu generalizace do modelu. Diskrétní varianta je definována:

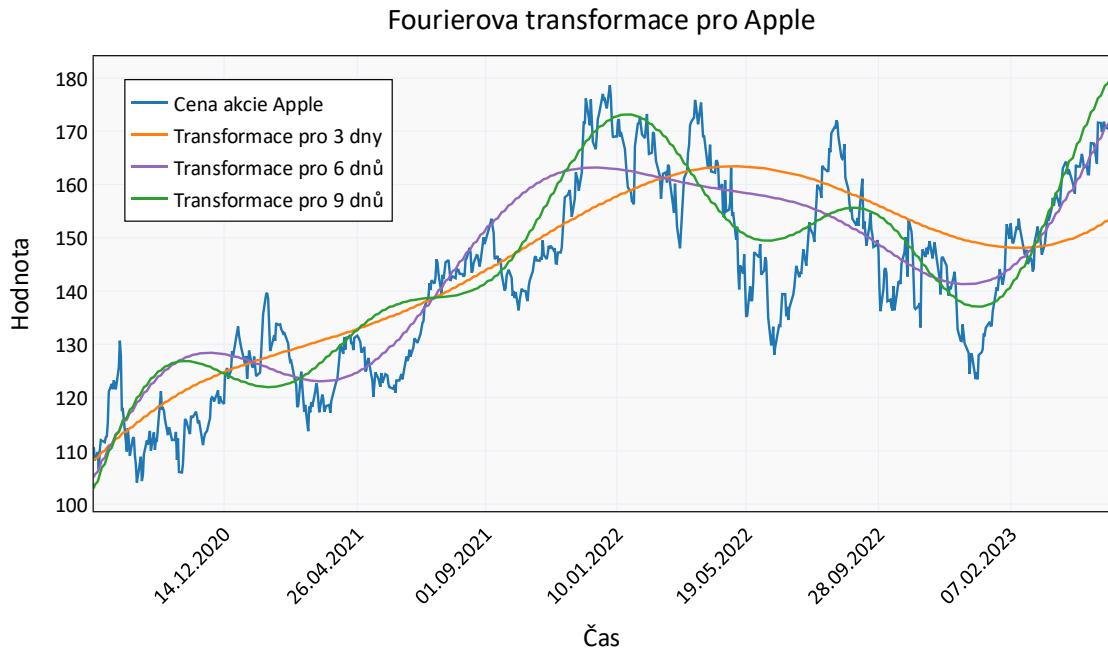
$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi\frac{kn}{N}}, \quad (28)$$

kde $x(n)$ označuje časovou řadu, n je časový index, N zachycuje celkovou velikost a k je index frekvenčního koše. K ní pak existuje inverzní verze definována následovně:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{i2\pi\frac{kn}{N}}, \quad (29)$$

kde $X(k)$ je Fourierův koeficient pro frekvenci koše k , i je imaginární jednotka a N je velikost. Všechny datasety obsahují transformace zpožděné o 3, 6 a 9 dní. Všechny jsou vypočítané z ceny na konci dne dané predikované komodity [37].

Nevýhodou jejich užívání je potřeba počítat transformace na velkém období, jelikož mají tendenci zkreslovat data na začátku a konci časových řad. Je tedy potřeba i pro predikce dalšího dne použít dostatečně dlouhý horizont minulých dat. Na obrázku 26 je zobrazen graf Fourierových transformací pro akcie Apple.



Obrázek 26: Graf Fourierových transformací počítaných s různými zpožděními pro cenu akcií Apple

4.3 Investorova pozornost

Investor při rozhodování o druhu investice využívá velké množství zdrojů. Typickými příklady jsou různé ekonomické ukazatele, jako jsou účetní uzávěrky či vliv ceny akcií, až po složitější nástroje, jako je fundamentální analýza. Často může být ukazatelem samotný chod firmy (nová větší zakázka, nový produkt či produktová řada). S rozšířením internetu a tím i zlepšením dostupnosti dat se důležitým aspektem stávají sociální sítě či zpravodajské weby (zprávy z novin měly vliv již dříve, díky internetu je ale dostupnost informací jednodušší a také investování se ulehčilo). To se nejčastěji zahrnuje pomocí sentimentální analýzy (hodnocení sentimentu textu zprávy/nadpisu/popisku). Dnes se proto využívají různé modely tohoto typu. Tato práce využívá model *FinBERT*, specializovaný na finanční zprávy a texty. Model vrací 3 hodnoty: pozitivní, neutrální a negativní skóre. Z nich se pak vypočítávají skóre polarity (PS) a jistoty (CS):

$$PS = positive - negative, \quad (30)$$

$$CS = 1 - neutral, \quad (31)$$

kde *positive* je pozitivní skóre, *negative* je negativní skóre a *neutral* je neutrální skóre.

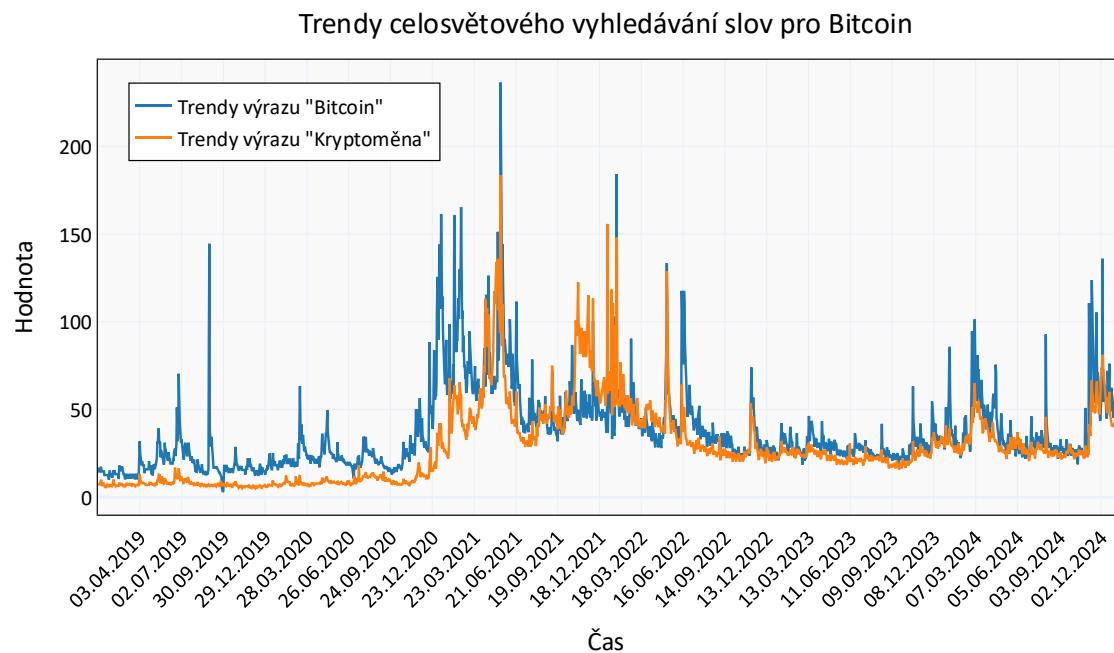
4.3.1 Trendy vyhledávání

Internetové vyhledávání může být ukazatelem velkých změn. Pokud například firma vydá nový produkt, úspěšný nebo ne, zvedne se vyhledávání jejich názvu či produktu (typickým příkladem je Apple a jeho výrobky jako jsou iPhone – zvedají vyhledávání, a i ceny akcií). Slouží tedy jako jeden z ukazatelů velikosti dané změny (i když nemusí vždy poukázat na to, zda je kladná či záporná).

Pro datasety využité v této práci je tato informace zastoupena z vyhledávání pomocí *Google Trends*. Hlavním aspektem není jen samotný název komodity, ale i jiné výrazy přidružené. Jmenovitě pro jednotlivé aktiva se jedná o:

- Apple – *Apple, iMac, iPhone*
- Bitcoin – *Bitcoin, Kryptoměny*
- Zlato – *Gold*

Trendy vyhledávání poskytuje Google volně ve formě dat v procentech skrze zadávané období. Bohužel denní data poskytuje pouze v kratším horizontu (9 měsíců) a je potřeba je spojit do celkových 5 let dat. Toho se docílí pomocí použitím překrývajících se oken a přepočet procent na základě stejných dnů. Hodnoty se následně vztahují na celé sledované období (proto hodnoty mohou mít i více jak 100 %). Na obrázku 27 je uveden graf výrazů pro Bitcoin.



Obrázek 27: Trendy vyhledávání skrze celé zahrnuté období pro predikce ceny Bitcoinu

4.3.2 Reddit

Sociální síť hraje v informačním poli obrovskou roli. S dostupností internetových služeb se zvedá i jejich důležitost. Dnes každá velká společnost má vlastní účty na nich, kde informuje o změnách či jiných věcech. Z interakcí jednotlivých uživatelů lze pak vysledovat určitá míra nálad. Díky jednoduššímu přístupu na trh pak mohou ovlivňovat trh i napřímo nákupem či prodejem daných akcií či opcí. Jako zástupce pro tento typ byl vybrán Reddit. Jedná se o sociální síť ve formě různých fór (*subreddits*), kde uživatelé přidávají příspěvky. Ty lze ohodnotit kladně (*upvote*) nebo záporně (*downvote*), komentovat a sdílet. V rámci sentimentální analýzy se pro výpočet využil nadpis, text, skóre a počet komentářů jednotlivých příspěvků. Souhrnná data pocházejí ze stránky academic torrents.com a autoři souhrnného balíku jsou (pod přezdívkami) stuck_in_the_matrix, Watchful1 a RaiderBDev [38].

Celkové skóre pro nadpis (TT) a text příspěvku (PT) se řídí:

$$TT = a \cdot PS_{title} \cdot CS_{title} + b \cdot \log(c \cdot comments_{norm} + d \cdot score_{norm}), \quad (32)$$

$$PT = a \cdot PS_{post} \cdot CS_{post} + b \cdot \log(c \cdot comments_{norm} + d \cdot score_{norm}), \quad (33)$$

kde PS_{title} , PS_{post} , CS_{title} a CS_{post} jsou skóre polarity a jistoty pro nadpis a text příspěvku, dále $comments_{norm}$ a $score_{norm}$ jsou normalizované počty komentářů a normalizované skóre Redditu (rozdíl mezi upvotes a downvotes). Posledními nepopsanými částmi jsou koeficienty a , b , c , d určující míru vlivu jednotlivých částí na celkové skóre. V rámci práce byly použity následující hodnoty:

- $a = 1,6$
- $b = 0,6$
- $c = 0,9$
- $d = 0,333$

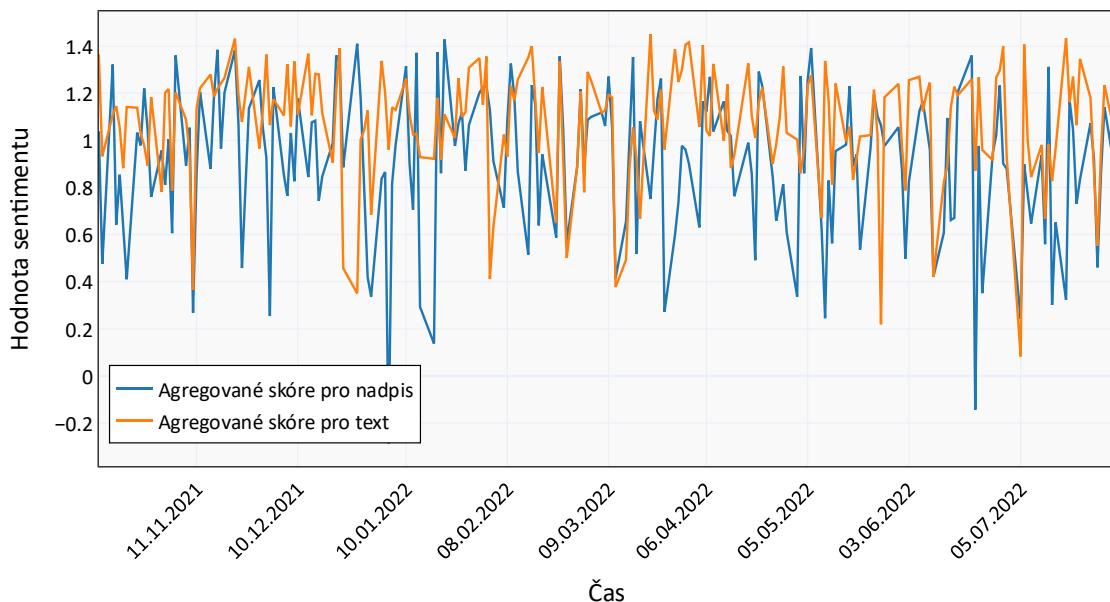
Z těchto hodnot vyplývá, že největší váhu má výpočet z výstupu z modelu *FinBERT*. Tyto hodnoty se vypočítají pro každý vybraný subreddit. Výsledná skóre se získají váženou sumou přes všechny subreddity, přičemž váhy jsou normalizované velikosti subredditů (hodnoty lze vidět v tabulce 1).

Tabulka 1: Tabulka vah pro všechny užité subreddity

	Apple	Bitcoin	Zlato
WallStreetBets	0,5550	0,3640	0,5460
Investing	0,0896	0,0594	0,0891
Options	0,0376	0,0249	0,0374
PennyStocks	0,0607	0,0402	0,0603
Stocks	0,2572	0,1705	0,2557
Bitcoin		0,1513	
CryptoCurrency		0,1897	
WallStreetSilver			0,0115

Dataset pro predikce cen akcií Apple zahrnuje informace z 5 subredditů. Jedná se o *WallStreetBets*, *Investing*, *Options*, *PennyStocks* a *Stocks*. *WallStreetBets* je zaměřeno na diskuse o akciích a opcích všeho druhu. Velmi se proslavilo zapojením do umělého vzniku nuceného výkupu (obchodníci spekulující na pokles ceny, tzv. *short*, jsou donuceni uzavřít pozice nákupem aktiva) akcií *GameStop*. *Investing* má hlavní orientaci na investice na kapitálových trzích. Oproti předešlému se snaží o vyšší standard diskuse, jedná se častěji o serióznější příspěvky od různých rad, kurzů po sdílení informativních článků. *Subreddit Options*, jak z názvu vyplývá, je zaměřen na obchodování opcí. Opět se jedná o vážnější fórum s přísnějšími pravidly. *PennyStocks* je více uvolněné, příspěvky se často zabývají vtipnými texty či obrázky. Přesto díky své velikosti jsou zdrojem nálad ve společnosti. Posledním zdrojem pro predikce Apple je *subreddit Stocks*. Zde se příspěvky zabývají novými či neznámými cennými papíry, zprávami ze světa financí nebo edukačními videi a texty.

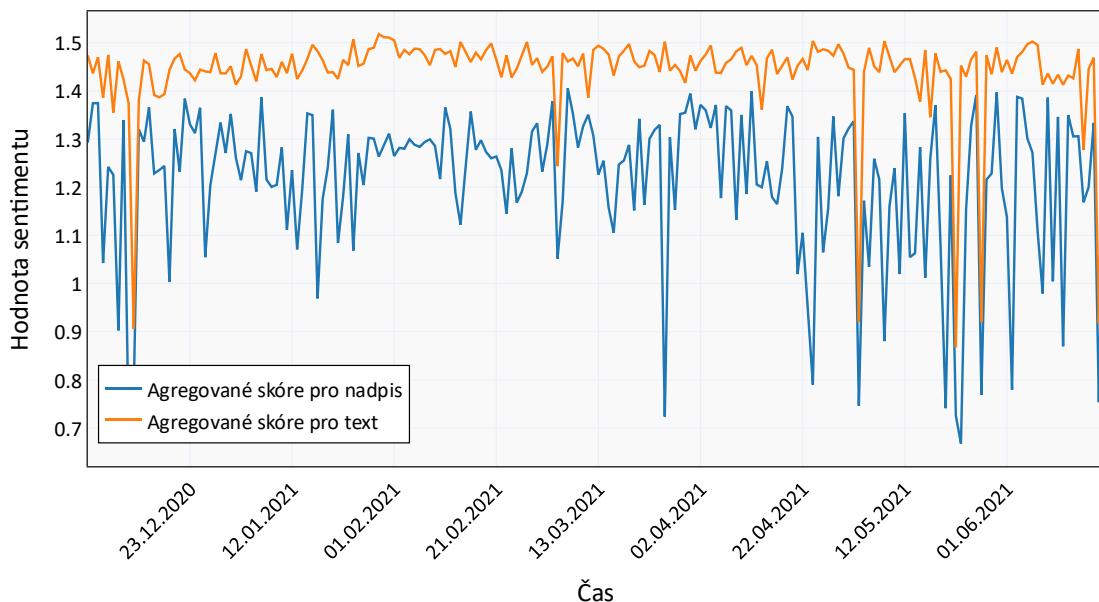
Agregované skóre sentimentu příspěvků z Redditu pro Apple



Obrázek 28: Graf korelace mezi sentimentem nadpisu a textu příspěvků, kratší období pro lepší čitelnost

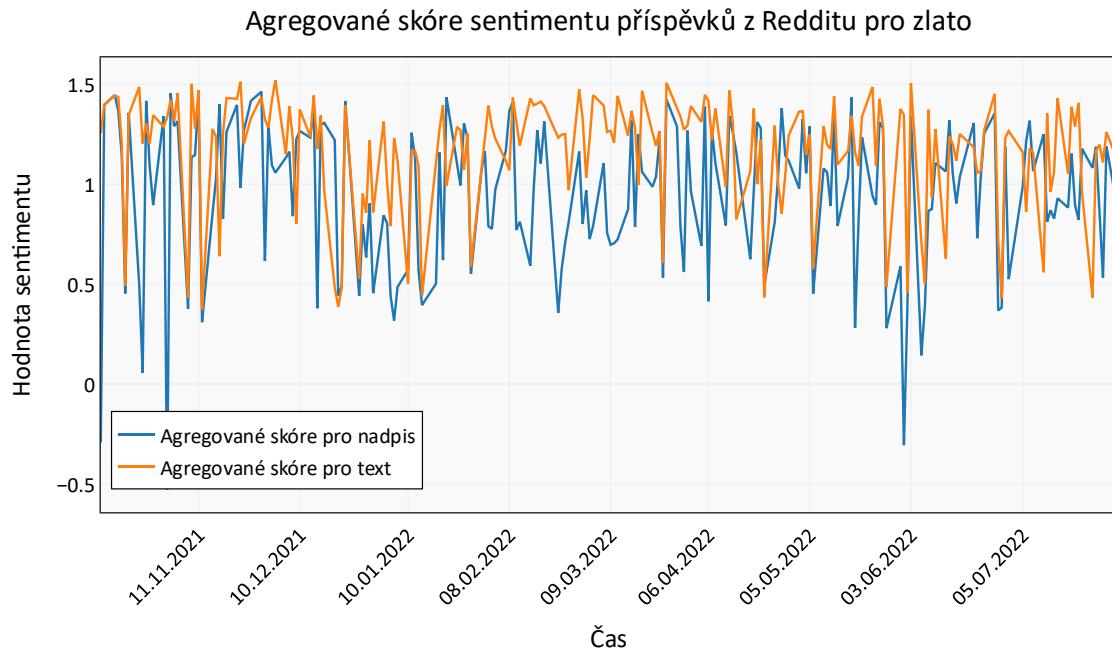
Předpověď ceny Bitcoinu, kromě již vyjmenovaných u Apple, obsahuje *subreddity Bitcoin* a *Cryptocurrency*. Z názvu prvního subredditu vyplývá jeho zaměření. Jedná se o kombinaci relevantních informací a různých anekdot na téma časté volatility ceny. *Cryptocurrency* pak zastřešuje diskusi nad tématem kryptoměn obecně, od zpráv a analýz až po zábavné postřehy.

Agregované skóre sentimentu příspěvků z Redditu pro Bitcoin



Obrázek 29: Graf vypočítaného sentimentu nadpisů a samotného textu příspěvku pro predikci Bitcoinu, kratší časový úsek pro lepší čitelnost

Analýza sociální sítě pro predikce ceny zlata, společně se subreddity pro Apple, zahrnuje i subreddit s názvem *WallStreetSilver*. Ten se spíše zaměřuje na stříbro ale i další drahé kovy zde mají v diskusi zastoupení.



Obrázek 30: Graf sentimentu nadpisů a textu příspěvků o zlatě

Příspěvky z vybraných subredditů pro určené predikce komodit byly vybrány na základě obsahu určitých slov či spojení. Jednalo se o celé názvy, zkratky, výrobky s ní spojené, zastřešující kategorie atp.

Dohromady se po filtrace jednalo o více než 1500000 různých příspěvků. Nejpočetnější byla skupina kolem kryptoměny Bitcoin (speciálně subreddity na něj zaměřené). Nejvíce příspěvku skrze všechny aktiva měl subreddit *Stocks*. Podrobné počty lze vidět v tabulce 2.

Tabulka 2: Souhrnná tabulka počtu příspěvků pro jednotlivá aktiva a subreddity

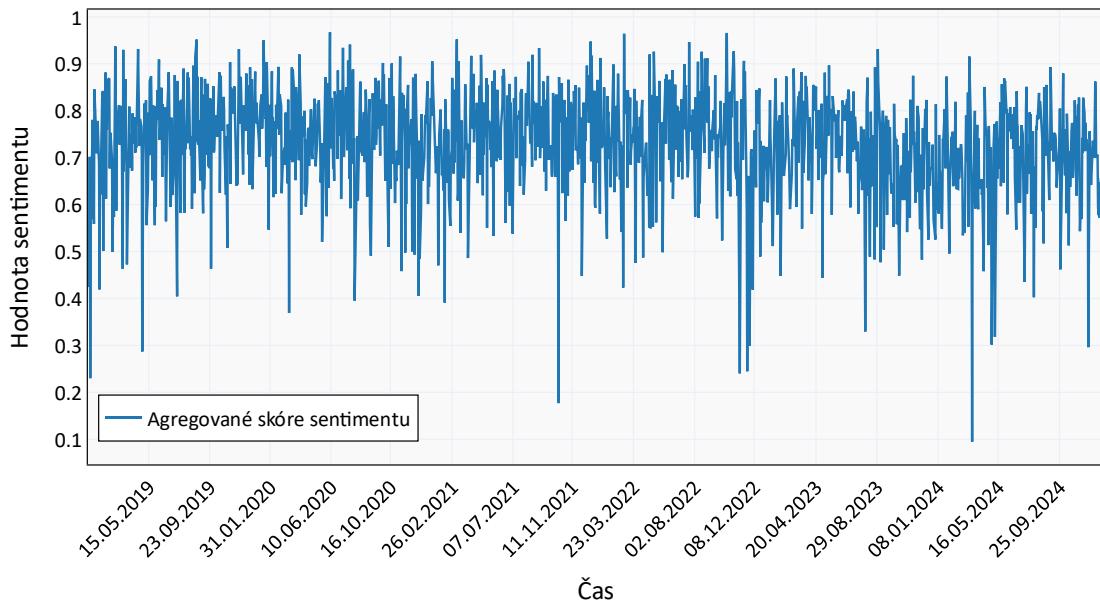
	Apple	Bitcoin	Zlato
WallStreetBets	15714	17090	9921
Investing	3266	7737	3982
Options	2353	1214	725
PennyStocks	1060	3736	3183
Stocks	8468	208260	305101
Bitcoin		315084	
CryptoCurrency		546152	
WallStreetSilver			47842
Celkový počet	30861	1099273	370754

4.3.3 Novinové titulky

Zprávy ze světa byly vždy silnou proměnnou v rámci investorovy pozornosti, dávno před vznikem internetu. V dnešním informačním prostoru se poptávka po kvalitních a pravdivých informacích nezmenšuje.

Data v této práci zahrnují více jak 6000 různých zdrojů získaných z *RSS Google News*. Dohromady se jednalo o zhruba 250000 článků. Titulky článků byly zpracovány obdobně jako text u Reddit příspěvků a bylo u nich vypočteno skóre polarity a jistoty. Skóre sentimentu daného titulku je pak dáno součinem jistoty a polarity. Vážená suma byla použita na agregování zpráv přes každý den. Zpravidajské kanály byly rozděleny do 3 kategorií na základě skóre získaného z LLM modelů (Claude a jiné). Tyto 3 kategorie měly hodnoty vah nastaveny na 0,3, 0,5 a 1. Určení, jaká zpráva patří do kategorie určité predikce, probíhalo obdobně jako u příspěvků z Redditu, na základě vybraných slov a spojení (nebo slov a spojení, která nesměla obsahovat).

Agregované skóre sentimentu novinových nadpisů pro Apple



Obrázek 31: Graf skóre sentimentu pro nadpisy novinových článků související s firmou Apple

5 Modely a jejich výsledky

Jak již bylo zmíněno výše, modely jsou tvořeny na predikce pro 3 různá aktiva. Data jsou před procesem učení upravena pomocí *MinMax* škálování od -1 do 1. Dále jsou rozděleny do 3 částí: trénovací, validovací a testovací. Trénovací sada se přímo podílí na učení modelu, tedy na úpravě vah a parametrů během trénovacího procesu. Validační sada slouží k vyhodnocení chyby modelu v průběhu trénování. Na základě validační chyby se zastavuje trénování dříve a díky tomu se zabraňuje přeúčení. Validační sada se také používá pro výběr nejlepší verze modelu nebo ladění hyperparametrů. Testovací sada slouží k finálnímu zhodnocení přesnosti modelu. Měla by být co nejvíce nezávislá na trénovacích i validačních datech, aby poskytla objektivní odhad toho, jak si model povede na zcela nových datech. Zároveň jsou členěny do tzv. oken (do modelu data nevstupují jako celek, ale jako kratší časový úsek 20 kroků). Tyto vzorky se pak rozdělí do dávek (*batch*) po 32. Toto členění mělo tvar 65 %/20 %/15 % pro Apple a zlato, 65 %/15 %/15 % pro Bitcoin. Predikce byly tvořeny na klouzavý průměr konečných cen (odhad trendu ceny) a na konečnou cenu dalšího dne. Data pro modely prošla filtrací vlastností pomocí kombinace *ANOVA* a *RFE* metody (nejdříve prošla data *ANOVA* a poté pokračovala do *RFE*).

5.1 Predikce trendu

Předpověď trendu je jedním ze způsobů predikcí na finančních trzích, zvláště pokud se jedná o zaměření na delší horizont investování (měsíce, roky). V této části jsou vyjmenovány výsledky pro tento typ modelů. Všechny používají jako cíl problému hodnoty pohyblivého průměru konečné ceny. Období pro transformaci jsou 7, 21 a 50 dní.

5.1.1 Apple

V tabulce 3 lze vidět výsledky pro jednotlivá okna průměrů. Nejlepší, co se týká metrik souvisejících s trendem dat, je model pro hodnotu 21. Pouze v přesnosti ho překonává model pro nejdélší období. Nejkratší má sice nejlepší hodnoty chyb, ale v ostatních metrikách ztrácí, nejvíce u citlivosti a F1 skóre.

Tabulka 3: Výsledky metrik pro předpověď trendu akcie Apple

	Okno 7			Okno 21			Okno 50		
	Trénování	Validace	Testovaní	Trénování	Validace	Testovaní	Trénování	Validace	Testovaní
MSE	20,9460	11,0052	38,0723	37,5099	32,5522	47,2994	28,6203	38,0977	56,4671
RMSE	4,5767	3,3174	6,1703	6,1245	5,7055	6,8775	5,3498	6,1723	7,5145
ACC	0,6025	0,5864	0,5656	0,5594	0,5802	0,5799	0,5295	0,5664	0,5581
PREC	0,7205	0,6748	0,6460	0,7080	0,7134	0,7025	0,6805	0,7134	0,7131
REC	0,6040	0,6145	0,5659	0,5646	0,5895	0,6028	0,5351	0,5864	0,5918
F1	0,6571	0,6433	0,6033	0,6282	0,6455	0,6489	0,5991	0,6437	0,6468
TRCOR	0,1972	0,1548	0,1342	0,1082	0,1461	0,1361	0,0511	0,1067	0,0724

5.1.2 Bitcoin

V tabulce 4 jsou zobrazeny výsledky pro předpověď trendu Bitcoinu. Okno velikosti 50 má nejmenší chybu, a to téměř o polovinu. V metrikách přesnosti trendu si nevede také špatně, avšak model s obdobím 21 předpovídá trend lépe, zejména v rámci citlivosti, F1 skóre a trendové korelace.

Tabulka 4: Výsledky metrik trendu pro predikce ceny Bitcoinu

	Okno 7			Okno 21			Okno 50		
	Trénování	Validace	Testovaní	Trénování	Validace	Testovaní	Trénování	Validace	Testovaní
MSE	23633787	1290666	32172496	47928148	2947916	27080658	6946916	1579206	7739469
RMSE	4861,459	1136,075	5672,080	6923,016	1716,950	5203,908	2635,700	1256,665	2781,990
ACC	0,580	0,579	0,670	0,598	0,595	0,677	0,638	0,637	0,615
PREC	0,611	0,621	0,718	0,668	0,649	0,763	0,690	0,773	0,735
REC	0,577	0,607	0,690	0,593	0,613	0,711	0,642	0,645	0,651
F1	0,594	0,614	0,704	0,629	0,631	0,736	0,665	0,703	0,690
TRCOR	0,160	0,151	0,331	0,195	0,184	0,323	0,274	0,254	0,189

5.1.3 Zlato

Tabulka 5 obsahuje výsledky jednotlivých metrik pro všechny modely trendů ceny zlata. Prakticky ve všech metrikách vyšel nejlépe model s 21 dny. Výjimkou je přesnost, kterou zvládl lépe model s nejdelším modelem. Model s obdobím 7 skončil s největší chybou, ale metriky předpovědi trendu jsou blízko hodnotám středního modelu.

Tabulka 5: Výsledky metrik pro predikce trendu ceny zlata

	Okno 7			Okno 21			Okno 50		
	Trénování	Validace	Testovaní	Trénování	Validace	Testovaní	Trénování	Validace	Testovaní
MSE	8548,05	4214,45	5052,79	2774,22	671,66	1867,32	652,97	717,55	2317,26
RMSE	92,4557	64,9188	71,0830	52,6709	25,9165	43,2124	25,5533	26,7871	48,1379
ACC	0,5739	0,6606	0,6290	0,6059	0,5708	0,6438	0,5518	0,5442	0,5581
PREC	0,6304	0,6909	0,7339	0,6406	0,6207	0,8480	0,6312	0,6126	0,9098
REC	0,5949	0,6496	0,6500	0,6198	0,5902	0,6424	0,5646	0,5528	0,5692
F1	0,6121	0,6696	0,6894	0,6300	0,6050	0,7310	0,5961	0,5812	0,7003
TRCOR	0,1406	0,3221	0,2356	0,2087	0,1359	0,2530	0,0969	0,0846	0,0113

5.2 Predikce ceny

V této části jsou výsledky modelů pro predikci ceny komodity následujícího dne. Pro evaluaci byly vybrány metriky: střední absolutní chyba, střední kvadratická chyba, směrodatná odchylka chyb, střední absolutní procentuální chyba a koeficient determinace (R^2).

5.2.1 Apple

Data pro predikce Apple po výběru vlastností obsahuje 64 finančních vlastností, 17 zpožděných proměnných a 6 vlastností sentimentu. Hyperparametry nejlepšího modelu byly vybrány na základě testování různých kombinací. Ty byly zároveň vybrány na základě vlastností modelu. Nejlepší model měl hyperparametry následující:

- Vstupní dimenze: 128
- Počet hlav: 8
- Míra odpadnutí: 0,1
- Míra vyhlazení hodnoty validace při učení: 0,04
- Míra učení: 0,0005

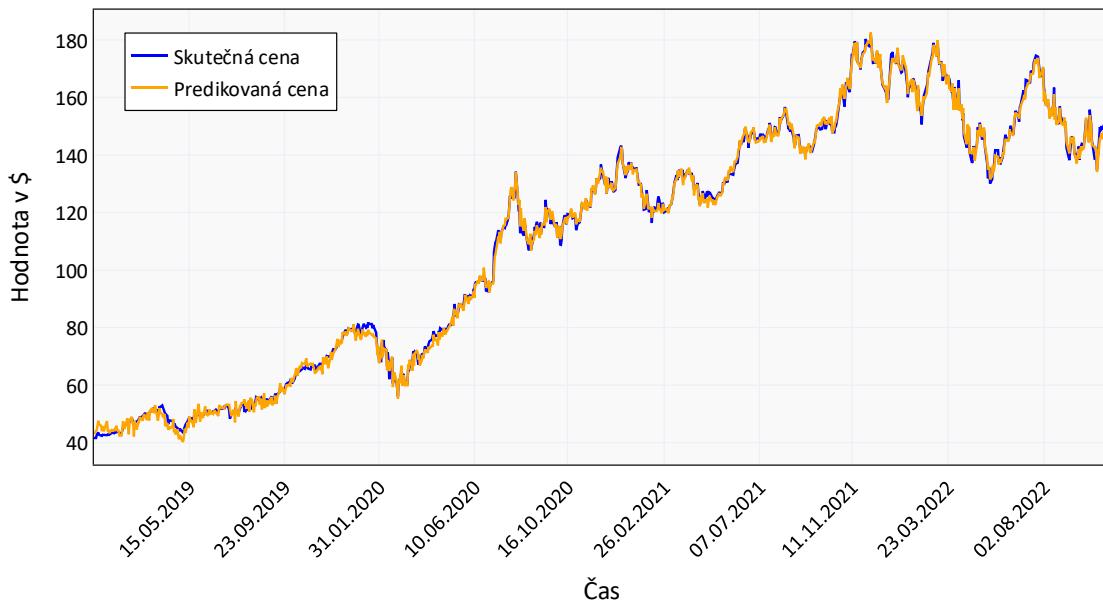
Jednalo se o nejjednodušší použitý model, přičemž hodnoty vstupní dimenze (128) a počtu hlav (8) vycházejí z osvědčených nastavení běžně využívaných u menších datových sad. Stejně tak míra odpadnutí (0,1) patří mezi konzervativní volby, které pomáhají omezit přeúčení, aniž by negativně ovlivňovaly schopnost modelu učit se z dat. Míra vyhlazení validační chyby našla využití především kvůli tendenci architektury predikovat více volatilní výstupy (např. jednoduchý enkodérový model na stejných datech je v tomto ohledu méně reaktivní). Na rozdíl od predikce ceny Bitcoinu se zde volatilita projevovala méně výrazně, a proto postačovala mírná hodnota vyhlazení. Celkové zhodnocení všech fází pomocí metrik lze nalézt v tabulce 6 níže.

Tabulka 6: Výsledky metrik přesnosti pro jednotlivé části tvorby a testování modelu pro předpověď ceny Apple

	Trénování	Validace	Testování
MAE	1,2724	2,0742	3,5303
MSE	2,5774	6,6621	23,8237
RMSE	1,6054	2,5811	4,8810
MAPE	0,0145	0,0121	0,0162
R2	0,9986	0,9816	0,9636

Při trénování dosahoval model velmi dobrých výsledků. Střední kvadratická chyba je 1,61 a koeficient determinace dosahuje 0,9986. Model se tedy dobře naučil vlastnosti trénovacích dat. Potvrzení pak lze vidět na obrázku 32, kde model dobře opisuje data. Jedinou větší odchylku lze pozorovat na začátku roku 2020, kde při trénování lehce podhodnotil cenovou hladinu.

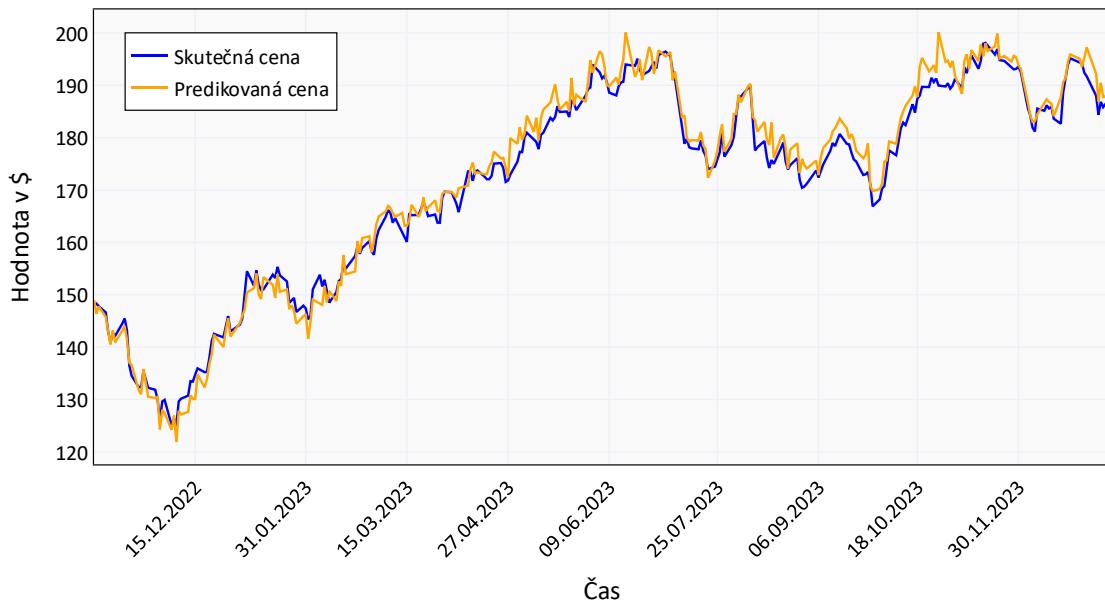
Predikce ceny akcie Apple - Trénovací data



Obrázek 32: Graf výsledků modelu predikcí Apple na trénovacích datech

Na validační části dat model produkuje výsledky s chybou 2,58. Obecně se chyba i v ostatních metrikách pohybuje blízko té trénovací. Na obrázku 33 predikce už neopisuje křivku tak přesně, ale stále se jedná o malé odchylky. Velkou část z nich tvoří spíše nadhodnocení změny, směr bývá správný.

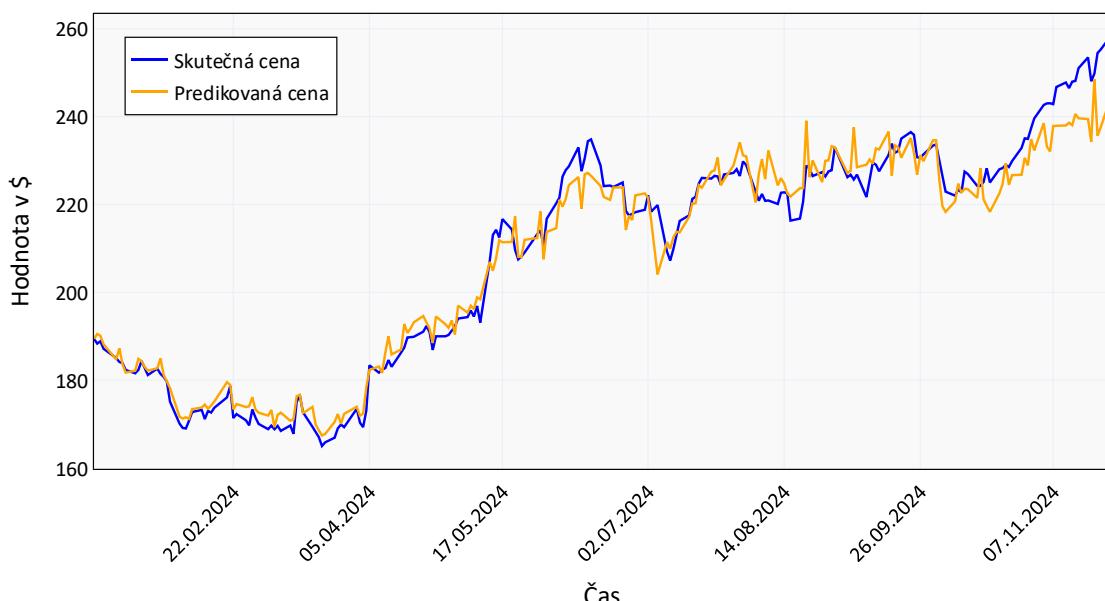
Predikce ceny akcie Apple - Validační data



Obrázek 33: Graf validace předpovědi akcie Apple

Poslední část datasetu je věnována samotnému testování. Střední kvadratická odchylka má zde hodnotu 4,88. Tím pádem model neztrácí schopnost předpovídат ani na neznámých datech. Z obrázku 34 je zjevné, že hlavně v první třetině si model vede skvěle, největší chyba nastane kolem června roku 2024. Zde nejdříve podhodnotil růst (směr i vrchol odhadl správně) a předpověděl pád ceny o několik dnů dříve v červnu roku 2024. To samo o sobě zvýší chybu numerickou, ale pro investice by se jednalo o nepatrný problém, jelikož poslední vrchol před ním odhadl správně. Na konci roku se pak cenová hladina skutečné ceny zvyšuje rychleji, než je predikce modelu, což mohlo být způsobeno náladou ve společnosti způsobenou volbou prezidenta USA.

Predikce ceny akcie Apple - Testovací data



Obrázek 34: Graf z testovací části procesu predikce Apple

5.2.2 Bitcoin

Dataset pro predikce Bitcoinu obsahuje 83 různých finančních ukazatelů, opět 17 zpožděných proměnných a 5 vlastností sentimentu trhu. Zvolený model byl navržen s ohledem na vyšší komplexitu dat a zvýšenou volatilitu této komodity, což se promítlo i do nastavení hyperparametrů:

- Vstupní dimenze: 256
- Počet hlav: 8
- Míra odpadnutí: 0,1
- Míra vyhlazení hodnoty validace při učení: 0,1
- Míra učení: 0,0002

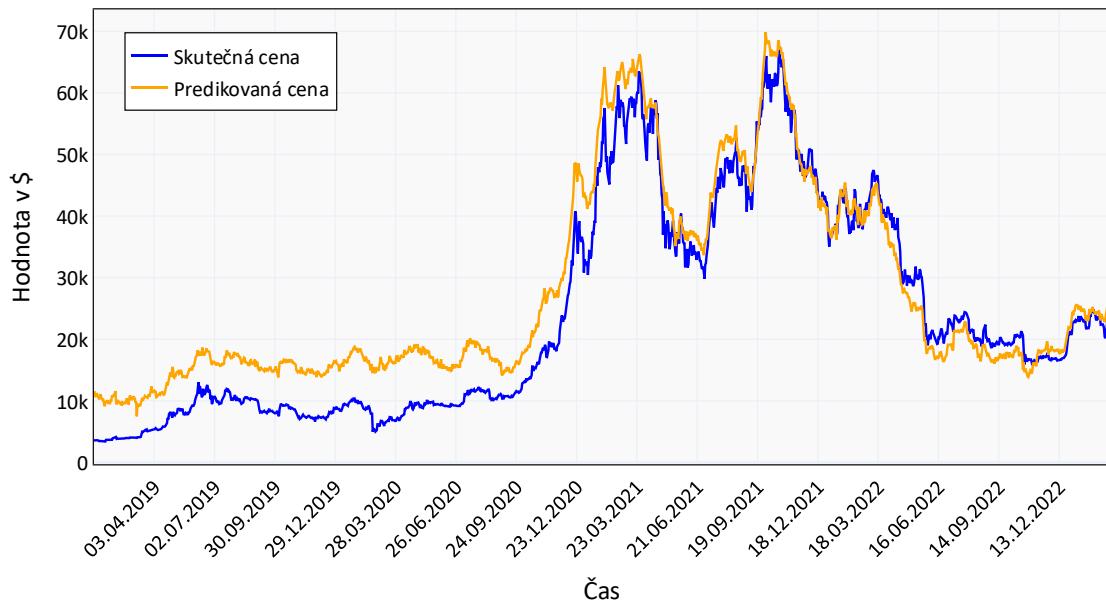
Vyšší dimenze zde reflektuje i větší složitost dat (oproti Apple predikci). Počet hlav zde odpovídá standardní hodnotě hlavně díky snížené míře učení, v této konfiguraci docházelo k nejmenšímu přeúčení modelu. Míra vyhlazení je tu největší ze všech použitých modelů, což je způsobenou velkou náchylností modelu na změny. Jedná se taky o potlačení vlivu validační části dat, jelikož její rozdelení se velmi lišilo od trénovací části datasetu. Zhodnocení modelu v rámci všech fází učení zobrazuje tabulka 7.

Tabulka 7: Výsledky metrik přesnosti pro předpověď Bitcoinu pro všechny fáze tvorby modelu

	Trénování	Validace	Testování
MAE	4894,5155	1282,6475	2847,5857
MSE	31698737,6118	2294117,5645	13909054,3165
RMSE	5630,1632	1514,6345	3729,4845
MAPE	0,4525	0,0411	0,0418
R2	0,8913	0,9412	0,9174

Trénink zde dosahuje chyby 5630, tedy o něco vyšší. V grafu na obrázku 36 je zjevné proč. Model změny v ceně předpovídá správně, ale cenovou hladinu v první polovině nadhodnocuje. Jedná se však o data před rokem 2021, kdy cena byla poměrně nízká (v porovnání s dalšími roky). Zároveň i informace ze sentimentální analýzy jsou v tomto období méně rozsáhlé. V druhé polovině se predikce již poměrně drží skutečné ceny.

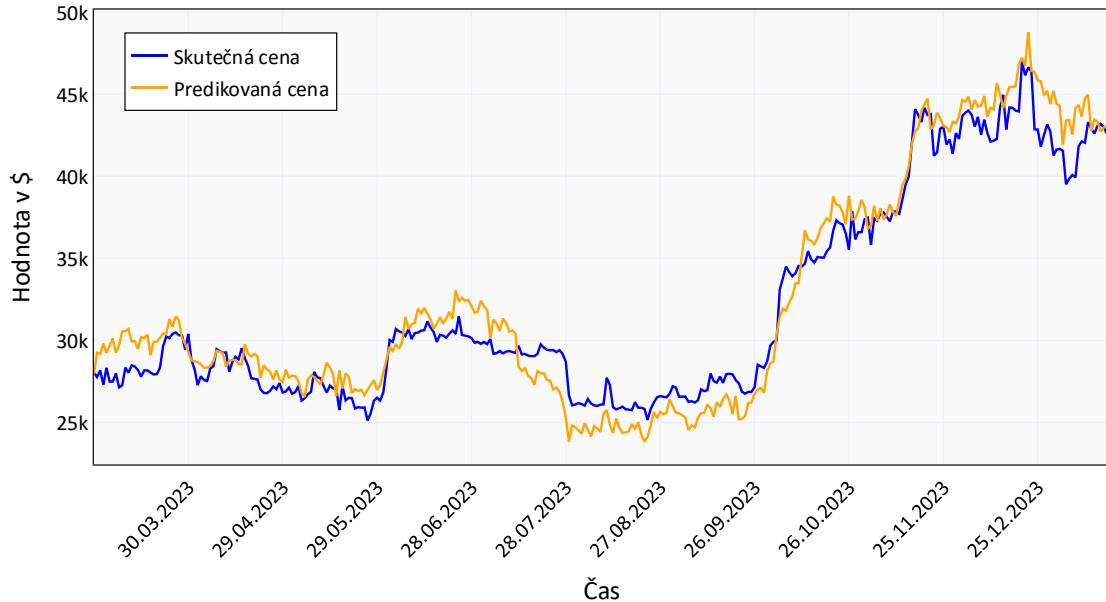
Predikce ceny kryptoměny Bitcoin - Trénovací data



Obrázek 36: Graf predikce na testovacích datech pro kryptoměnu Bitcoin

Avšak validace má již o dost nižší chybu. V rámci střední kvadratické odchylky se hodnota snížila na 1514. Model mohl upřednostnit informace z validační části dat, s největší pravděpodobností to však souvisí s distribucí dat. Validační část byla z pohledu vývoje ceny „lehčí“ pro predikce než trénovací. To lze vidět i na grafu hodnot na obrázku 35.

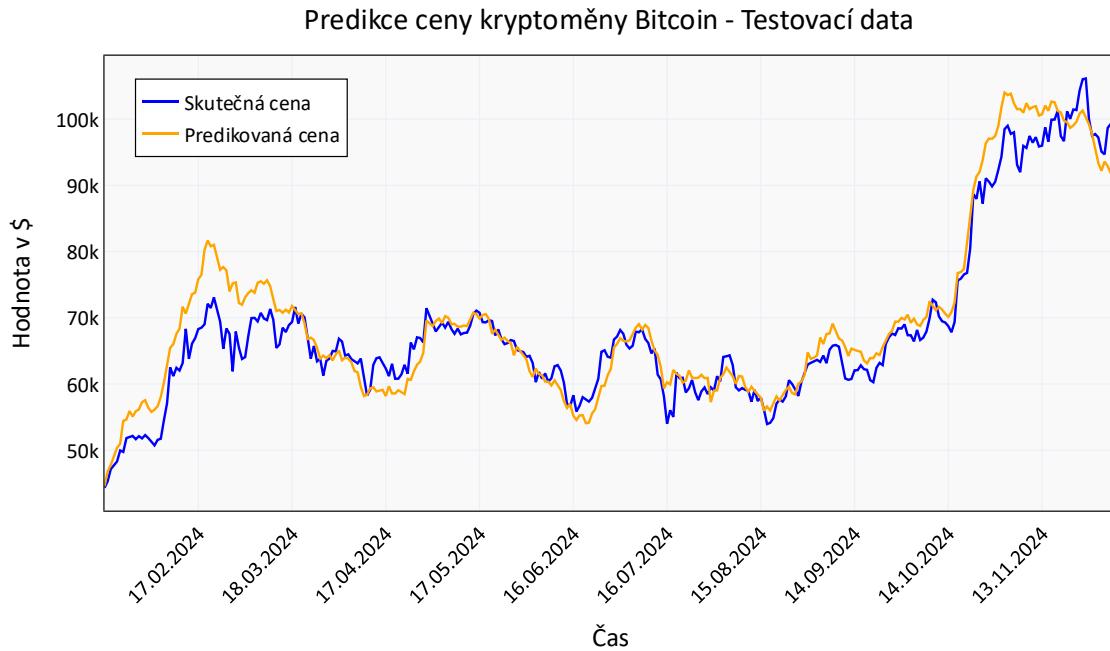
Predikce ceny kryptoměny Bitcoin - Validační data



Obrázek 35: Graf modelu na validačních datech pro dataset Bitcoin

Testování skončilo na hodnotě 3729. Koeficient determinace je 0,9174. Obecně na neznámých datech model dobře predikuje velké „vlny“, myšleno velké zvýšení a pokles ceny. Velké chyby spíše způsobuje nadhodnocení velikosti těchto změn (například na začátku roku 2024).

Posunutí predikce, které je obvyklé u dat s velkou volatilitou, se zde vyskytuje minimálně, znatelné je to pouze v růstu mezi červnem a červencem (lze vidět na obrázku 37).



Obrázek 37: Graf testování modelu předpovědi ceny Bitcoinu

5.2.3 Zlato

Použitá data obsahují celkové 111 vstupních proměnných, konkrétně se jedná o 90 finančních ukazatelů, 17 zpozděných proměnných a 4 vlastnosti sentimentu. S ohledem na povahu dat a testováním různých kombinací skončil nejlepší model s následujícími hyperparametry:

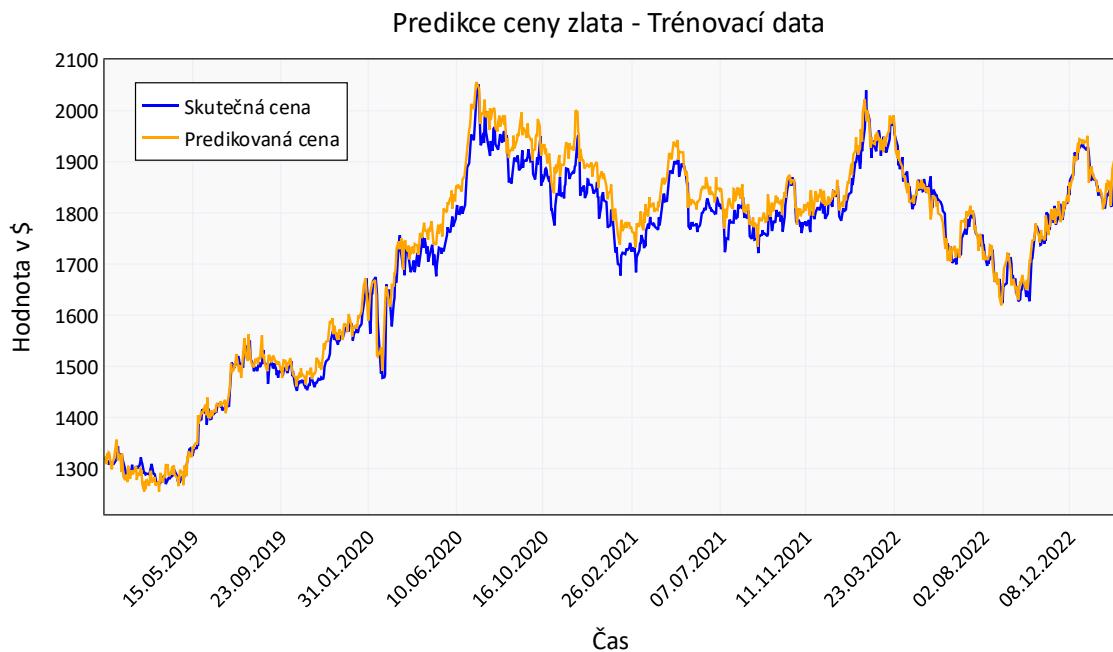
- Vstupní dimenze: 128
- Počet hlav: 16
- Míra odpadnutí: 0,1
- Míra vyhlazení hodnoty validace při učení: 0
- Míra učení: 0,0002

Největším rozdílem oproti předešlým modelům je zvětšený počet hlav. Ten byl vybrán z důvodu potřeby zaměření se na jemnější změny v rámci ceny. Společně s nízkou hodnotou učení se tato kombinace projevila jako kompromisní volba zabraňující přeúčení. Zároveň zde nebylo aplikováno vyhlazení v průběhu trénování (lze vidět na výsledcích větší tendenci vytvářet ostřejší změny ceny). Validaci část dat pro zlato má rozmanitější rozdělení a upozadění přijímání změn pak způsobovalo přílišné vyhlazený predikce. Výsledky metrik jsou uvedeny v tabulce 8.

Tabulka 8: Výsledky metrik přesnosti pro model predikce ceny zlata skrze všechny fáze

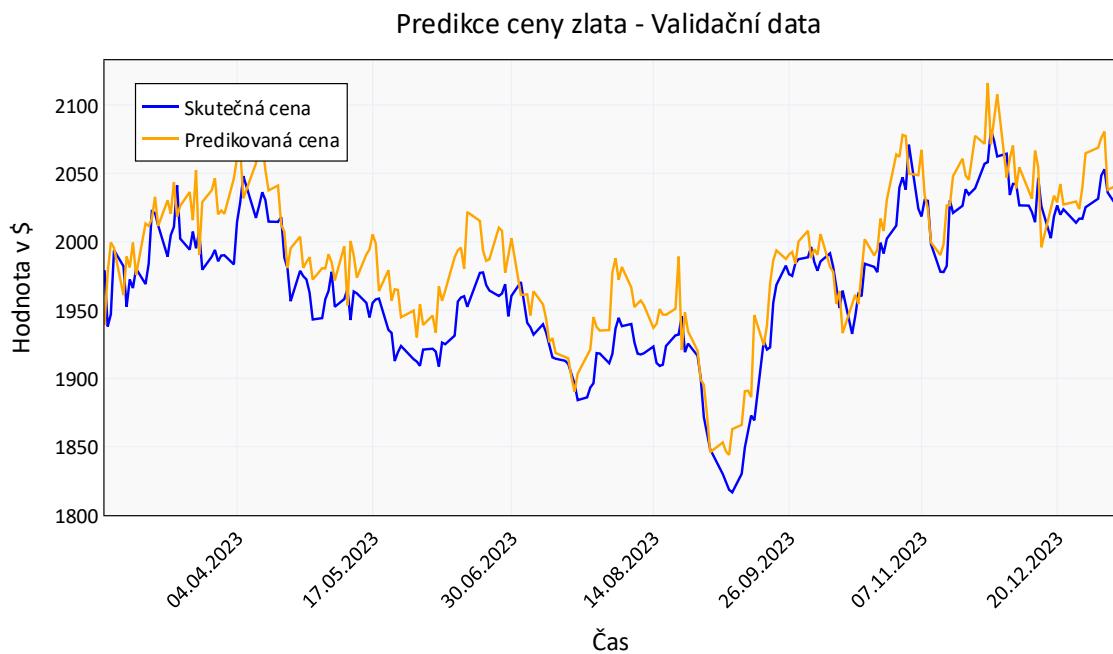
	Trénování	Validace	Testování
MAE	27,17	25,38	48,85
MSE	1182,68	908,75	3248,70
RMSE	34,39	30,15	57,00
MAPE	0,02	0,01	0,02
R2	0,97	0,67	0,92

V rámci trénování měl model velmi dobrý koeficient determinace – 0,97. To poukazuje na dobrou míru učení z trénovacích dat. To potvrzuje i graf z obrázku 38. Většinu chyby tvoří lehké nadhodnocení v rámci jednoho roku v polovině testovacích dat.



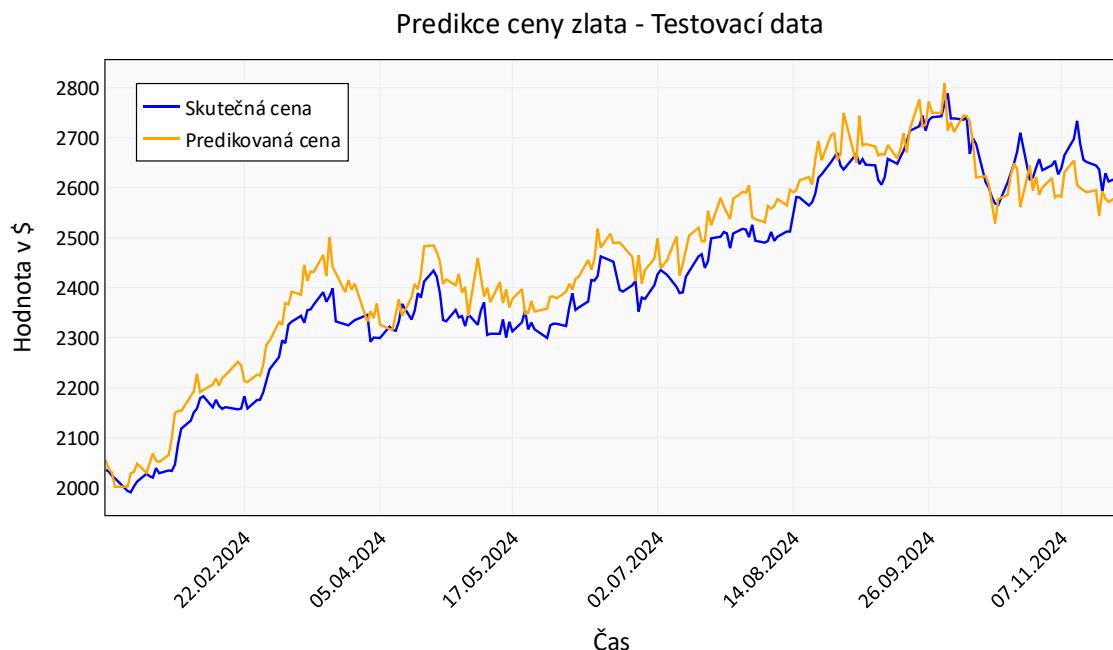
Obrázek 38: Graf modelu na testovacích datech pro predikci ceny zlata

Validace modelu měla menší chybu než trénovací část, přesněji 30,15 (u trénovací části je chyba 34,49). Koeficient determinace však skončil na nízké hodnotě, a to 0,67. Data zde opisují celkový trend je zde větší volatilita v predikcích, hlavně v porovnání s testováním.



Obrázek 39: Graf validace modelu předpovědi ceny zlata

Část testovací dosáhla střední kvadratické chyby 57 a koeficient determinace 0,92 (výrazné zlepšení oproti validaci). Trend model předpovídá velmi dobře. Nejvíce se odchyluje od skutečných dat ve výši dané změny (především lze vidět v březnu, kde předpověď růst, ale nadhodnocený). Oproti dřívějším modelům lépe zvládá i konec roku 2024 (i tak je zde vidět obdobně jako u Apple špatné vyhodnocení jednoho z trendů). To může být způsobeno konzervativnějším charakterem vývoje ceny zlata oproti Bitcoinu či Applu.



Obrázek 40: Graf vývoje predikce na testovacích datech pro cenu zlata

5.3 Porovnání s jinými modely

Porovnání se zaměří na práce na téma predikcí na finančních trzích se zahrnutím predikce ceny akcií Applu. Cílem je posoudit výkonnost navrhovaného řešení se současnými často používanými modely.

Typickým modelem pro predikce jsou různé druhy rekurentních sítí, převážně LSTM. Tento model používá práce *Stock price prediction: comparison of different moving average techniques using deep learning model*. Autoři v ní porovnávají predikci využitím LSTM s více fundamentální metodou pomocí klouzavých průměrů. Provedli porovnání s více typy předpovídaných akcií, všechny jako data použily hladiny jejich cen z trhu (myšleno Low, High, Open, Close) z let 2019 až 2022 (od května do května). Chybu RMSE pro predikci AAPL deklarují jako 6,44 se střední absolutní procentuální chybou 2,97 %. V porovnání s SMA a EMA modely pak dosahovali zhruba třetinového vylepšení [39].

Další prací využívající LSTM je *Advanced Stock Market Prediction Using Long Short-Term Memory Networks: A Comprehensive Deep Learning Framework* z roku 2025. Ta předpovídala ceny akcií Apple, Alphabet, Microsoft a Amazon pomocí jak finančních dat z burzy těchto společností, tak sentimentální analýzy novinových článků (Bloomberg a Reuters). Pokud se zaměříme na přesnost u predikce společnosti Apple, jejich model má střední kvadratickou odchylku 7,62 a střední absolutní procentuální chybu 2,72 %. Díky tomuto výsledku, jak sami

autoři uvádí, překonávají klasický model ARIMA (dnes se někdy využívá v kombinaci s ostatními modely) [40].

Práce *A Hybrid Convolutional Recurrent (CNN-GRU) Model for Stock Price Prediction* [41] využívá další z metod RNN. Ta prozkoumává hned několik hybridních modelů, jedná se o kombinace konvoluční sítě s rekurentní sítí, LSTM a GRU. Autoři predikovali více druhů akcií, jmenovitě Apple, TATA, Yahoo, Microsoft, Tesla a Uniqlo. Pro Apple mají rozsáhlý dataset, obsahující data mezi 1980 a 2021. Jako hlavní metriky pak využívají RMSE a R^2 . Nejlepším modelem jim vyšla kombinace CNN a GRU, která skončila s chybou 4,96 a koeficientem determinace 0,979. Model CNN a LSTM však neměl až tak rozdílné výsledky.

Poslední zmíněnou prací je práce *Enhancing stock price prediction using GANs and transformer-based attention mechanisms* [42]. Ta také využívá rozmanitá data v rámci investorovy pozornosti. Využívají finanční data více společností, různé technické indikátory a také sentimentální analýzu zpráv médií. Nejedná se však o průřez skrz titulky různých mediálních domů, ale jsou všechny ze zdroje se zaměřením na finanční zprávy (jsou zpracovávány pomocí modelu FinBERT). Samotné modelování ceny pak provádí GAN. Jejich model reprezentují různými metrikami, pro porovnání využijeme opět střední kvadratickou chybu, která činí 9,4235.

V porovnání s modelem navrženým v této práci, lze získat poměrně dobré vylepšení oproti LSTM nebo oproti práci využívající analogické složení dat (do kategorií myšleno, bylo více zaměřeno na finanční sektor), ale užívající model GAN. Srovnatelné výsledky vykazovalo použití hybridního modelu (koeficient determinace zde vychází lépe, chybí však další metriky, které by umožnily podrobnější zhodnocení vlastností spojeného modelu). Přesnější porovnání pak lze nalézt níže v tabulce 9.

Tabulka 9: Porovnání výsledků vybraných prací z posledních let, přeškrtnuté buňky představují chybějící výsledek pro danou metriku

	RMSE	MAPE	R^2
SMA	9,49	4,49 %	
EMA	8,1	3,89 %	
LSTM	6,44	2,97 %	
LSTM	7,62	2,72 %	
CNN-RNN	9,14		0,93
CNN-LSTM	6,71		0,962
CNN-GRU	4,96		0,979
GAN	9,42		
TFT (navr.)	4,88	1,21 %	0,963

Zároveň jsem se v rámci výzkumného úkolu zabýval predikcí několika různých aktiv pomocí GAN modelu, konkrétně kombinace GRU generátoru a CNN diskriminátoru. Předmětem predikce byly ceny akcií Apple, kryptoměny Bitcoin, opcí na zlato a akcií společnosti ČEZ. Model pracoval s finančními daty, zpožděnými indikátory a trendy vyhledávání.

Z hlediska celkové koncepce šlo o jednodušší přístup – zejména z pohledu vstupních dat. Z pohledu samotné architektury modelu pak záleží na konkrétní variantě TFT, se kterou je GAN

porovnáván (v této práci je použita zjednodušená verze modelu TFT, což činí porovnání relevantním, avšak je potřeba vzít v potaz i hyperparametry modelů pro správné porovnání).

Při predikci ceny akcií Apple dosáhl GAN model střední kvadratické chyby (RMSE) ve výši 6,51. V porovnání s výsledky modelu použitým v této práci představuje kombinace TFT a hlubší zapojení investorovy pozornosti výrazné zlepšení. Zároveň se u GAN modelu projevoval častý problém zpožděné reakce na cenové extrémy, kdy predikce zaznamenala posun oproti reálným hodnotám. Tento nedostatek se podařilo v aktuálním modelu výrazně omezit. Využití investorovy pozornosti přidává modelu informace o významných tržních změnách a architektura typu TFT díky své struktuře dokáže efektivněji zpracovat a promítnout je do výsledné predikce.

Závěr

Tato práce měla za cíl prozkoumat vliv investorovy pozornosti na cenu aktiv na finančních trzích pomocí transformerových modelů. Zahrnutá data obsahovala jak typická finanční data, tak i analýzu sentimentu veřejného prostoru. Konkrétně se jednalo o trendy vyhledávání slov, novinové články a příspěvky ze stránky Reddit. Na základě těchto dat byly vytvořeny modely pro predikci koncové ceny a jejího trendu.

Data v rámci investorovy analýzy zastřešují nejen různé finanční ukazatele, ale také se ukazuje, že nálada ve společnosti hraje rostoucí roli v konečných rozhodnutích. To souvisí nejen s jednodušší komunikací informací, ale i s jednodušším vstupem na trh, kdy pomocí různých služeb může investovat prakticky každý. Data vyjadřující sentiment jsou většinou obtížněji dostupná a je potřeba je více zpracovávat, aby byla vhodná pro použití (jedná se převážně o textové zprávy, které je potřeba kvantifikovat; dříve to byla často manuální práce, v současnosti spíše úkol modelů zpracování jazyka). Přesto je zakomponování různých vrstev dat investorovy pozornosti přínosným dodatkem, jak ukázaly výsledky této práce.

Transformerové modely použité pro demonstrování vlivu byly jednodušší verze architektury TFT, založené na architektuře dekodéru. Zaměření bylo na více druhů různých finančních komodit, konkrétně se jednalo o akcie, opce a kryptoměny. Každá z nich je spojena s různými vlastnostmi. Například kryptoměny jsou známé svou vysokou volatilitou, zatímco zlato má spíše mírnější výkyvy cen.

Předpovědi trendů pomocí klasifikačních modelů demonstrovaly schopnost zachytit směr pohybu u všech tří druhů aktiv. Obecně se nejlépe jevilo okno velikosti 21, které dostatečně zachycovalo minulá data. Nejlepším modelem byl model pro Bitcoin se skóre F1 0,736 a přesností 0,763. Zde se mohly projevit právě zmíněné informace z médií a sociálních sítí, které jsou u tohoto aktiva velice zásadní. U predikce ceny na další den opět modely zvládly předvést schopnost ji modelovat bez větších chyb. Obecně se u všech modelů nejobtížněji projevil konec roku 2024, tam se mohly projevovat jiné celospolečenské události, které přímo nesouvisely se sledovanými aktivy (nejpravděpodobněji se jedná o vliv voleb amerického prezidenta konaných v listopadu). Nejpřesnějším modelem dle metrik vyšel model pro ceny Applu, který na testovací sadě dat vykazoval střední kvadratickou chybu 4,88, střední absolutní procentuální chybu pouze 1,21 % a koeficient determinace 0,963. Tento výsledek byl porovnán s některými současnými pracemi na obdobné téma zaměřené převážně na obvyklé typy modelů. Navržený model dosahuje lepších výsledků než *LSTM*, a i vylepšení oproti *GAN* modelu používajícímu podobný typ dat. Hybridní modely s využitím kombinace více struktur měly výsledky u nejlepšího modelu srovnatelné. Avšak stojí za zmínu, že data, na kterých byl model stavěn, byla rozsáhlá (jednalo se o 40 let historie firmy Apple na finančních trzích) a jednalo se tedy o celkem velký model.

Přidaná data, která více prohlubují investorovu pozornost na trhu, se prokázala mít pozitivní efekt na predikci ceny. Společně s transformerovými modely vykazují lepší výsledky než současné klasické modely používající pouze finanční data dané komodity.

Vylepšením či pokračováním by mohlo být přidání dalších vlastností zahrnujících celospolečenské události. Jak bylo zmíněno výše, modely na ně reagovaly ne zcela přesně. Dále by přidání dalších sociálních sítí mohlo přinést nový rozměr, kam sentiment trhu směřuje (myšleno je například X.com, dříve známý jako Twitter.com, zde jsou však novější data těžko dostupná). Pokud by se více prohloubila míra predikce (délka předpovědního horizontu,

robustnost modelu), velkým přínosem pro vylepšení predikce by bylo přiblížení se původnímu modelu (například statické kovariáty mohou dodat určitý stabilizační prvek).

Zdroje

- [1] SARKER, Iqbal H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN computer science*, 2021, 2.6: 1-20.
- [2] AGGARWAL, Charu C., et al. *Neural networks and deep learning*. Cham: Springer, 2018.
- [3] ZHANG, Aston, et al. *Dive into deep learning*. Cambridge University Press, 2023.
- [4] VASWANI, Ashish, et al. Attention is all you need. *Advances in neural information processing systems*, 2017, 30.
- [5] BAHDANAU, Dzmitry. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [6] SHAZER, Noam. Fast transformer decoding: One write-head is all you need. *arXiv preprint arXiv:1911.02150*, 2019.
- [7] AINSLIE, Joshua, et al. Gqa: Training generalized multi-query transformer models from multi-head checkpoints. *arXiv preprint arXiv:2305.13245*, 2023.
- [8] DEVLIN, Jacob, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 2019. p. 4171-4186.
- [9] RADFORD, Alec, et al. Improving language understanding by generative pre-training. 2018.
- [10] BROWNLEE, Jason. *Probability for machine learning: Discover how to harness uncertainty with Python*. Machine Learning Mastery, 2019.
- [11] HASTIE, Trevor, et al. *The elements of statistical learning: data mining, inference, and prediction*. New York: springer, 2009.
- [12] RUDER, Sebastian. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.
- [13] LOSHCHILOV, Ilya; HUTTER, Frank. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [14] D'ANGELO, Francesco, et al. Why do we need weight decay in modern deep learning?. *Advances in Neural Information Processing Systems*, 2024, 37: 23191-23223.
- [15] LIM, Bryan, et al. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 2021, 37.4: 1748-1764.
- [16] RIZINSKI, Maryan, et al. Sentiment analysis in finance: From transformers back to explainable lexicons (xlex). *IEEE Access*, 2024, 12: 7170-7198.

- [17] JAHIN, Md Abrar, et al. A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. *Scientific Reports*, 2024, 14.1: 24882.
- [18] KONSTANTINIDIS, Thanos, et al. Finllama: Financial sentiment classification for algorithmic trading applications. *arXiv preprint arXiv:2403.12285*, 2024.
- [19] WANG, Chaoyang; LIU, Guangyu. From anomaly detection to classification with graph attention and transformer for multivariate time series. *Advanced Engineering Informatics*, 2024, 60: 102357.
- [20] KANG, Hyeongwon; KANG, Pilsung. Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism. *Knowledge-Based Systems*, 2024, 290: 111507.
- [21] LIU, Yong, et al. Timer: Generative pre-trained transformers are large time series models. *arXiv preprint arXiv:2402.02368*, 2024.
- [22] WANG, Yuxuan, et al. Timexer: Empowering transformers for time series forecasting with exogenous variables. *Advances in Neural Information Processing Systems*, 2024, 37: 469-498.
- [23] WOO, Gerald, et al. Unified training of universal time series forecasting transformers. 2024.
- [24] CHEN, Peng, et al. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. *arXiv preprint arXiv:2402.05956*, 2024.
- [25] DING, Qianggang, et al. Hierarchical multi-scale Gaussian transformer for stock movement prediction. In: *Ijcai*. 2020. p. 4640-4646.
- [26] ZHANG, Qiuyue, et al. Transformer-based attention network for stock movement prediction. *Expert Systems with Applications*, 2022, 202: 117239.
- [27] YAÑEZ, Camilo; KRISTJANPOLLER, Werner; MINUTOLO, Marcel C. Stock market index prediction using transformer neural network models and frequency decomposition. *Neural Computing and Applications*, 2024, 36.25: 15777-15797.
- [28] TAO, Zicheng; WU, Wei; WANG, Jianxin. Series decomposition Transformer with period-correlation for stock market index prediction. *Expert Systems with Applications*, 2024, 237: 121424.
- [29] REN, Shangsheng, et al. A novel hybrid model for stock price forecasting integrating Encoder Forest and Informer. *Expert Systems with Applications*, 2023, 234: 121080.
- [30] WANG, Shuzhen. A stock price prediction method based on BiLSTM and improved transformer. *IEEE Access*, 2023, 11: 104211-104223.
- [31] HARYONO, Agus Tri; SARNO, Riyanarto; SUNGKONO, Kelly Rossa. Transformer-gated recurrent unit method for predicting stock price based on news sentiments and technical indicators. *ieee Access*, 2023, 11: 77132-77146.

- [32] LEVY, S. a MONTEVIRGEN, K. *Apple Inc.* Online. Encyclopædia Britannica. 2025. Dostupné z: <https://www.britannica.com/money/Apple-Inc/2011-present-Tim-Cook-as-CEO-and-the-first-trillion-dollar-company>. [cit. 2025-07-06].
- [33] CHEN, J. *What is EMA? How to Use Exponential Moving Average With Formula.* Online. Investopedia. Dostupné z: <https://www.investopedia.com/terms/e/ema.asp>. [cit. 2025-07-06].
- [34] BOLLINGER, John. Using bollinger bands. *Stocks & Commodities*, 1992, 10.2: 47-51.
- [35] MONTERVIGEN, K. *MACD (moving average convergence divergence).* Online. Encyclopædia Britannica. Dostupné z: <https://www.britannica.com/money/macd-moving-average-convergence-divergence>. [cit. 2025-07-06].
- [36] HARI, Yulius; DEWI, Lily Puspa. *Forecasting system approach for stock trading with relative strength index and moving average indicator.* 2018. PhD Thesis. Petra Christian University.
- [37] SUTER, Bruce W. *Multirate and wavelet signal processing.* Elsevier, 1997.
- [38] STUCK_IN_THE_MATRIX; WATCHFUL1 a RAIDERSDEV. *Reddit comments/submissions 2005–06 to 2024–12.* Online. Academic torrents. [cit. 2025-07-06].
- [39] BILLAH, Md Masum, et al. Stock price prediction: comparison of different moving average techniques using deep learning model. *Neural Computing and Applications*, 2024, 36.11: 5861-5871.
- [40] CHAUDHARY, Rajneesh. Advanced Stock Market Prediction Using Long Short-Term Memory Networks: A Comprehensive Deep Learning Framework. *arXiv preprint arXiv:2505.05325*, 2025.
- [41] JAISWAL, Rashi; SINGH, Brijendra. A hybrid convolutional recurrent (CNN-GRU) model for stock price prediction. In: *2022 IEEE 11th international conference on communication systems and network technologies (CSNT)*. IEEE, 2022. p. 299-304.
- [42] LI, Siyi; XU, Sijie. Enhancing stock price prediction using GANs and transformer-based attention mechanisms. *Empirical Economics*, 2025, 68.1: 373-403.
- [43] GODOY, Daniel Voigt. *Deep Learning with PyTorch Step-by-step: A Beginner's Guide.* 2022.
- [44] MARWAX a MRMW. *Full GPT architecture.* Online. In: Wikipedia: the free encyclopedia. San Francisco (CA): Wikimedia Foundation, 2001-. Dostupné z: https://commons.wikimedia.org/wiki/File:Full_GPT_architecture.svg. [cit. 2025-08-03].